Investigation of glycosyltransferases from oat

Thomas LOUVEAU

A thesis submitted to the University of East Anglia for the degree of
Doctor of Philosophy

University of East Anglia
John Innes Centre
Norwich, the United Kingdoms
© October 2013

I

Abstract

Plants produce a diversity of secondary metabolites crucial for their survival into specific ecological niches. Many of these compounds are glycosides generated by the action of family one UDP-dependant glycosyltransferases (UGTs). Glycosylated products of UGTs are known to be essential for reproductive fitness, defence against pathogens, and signalling; UGTs also have a role in the detoxification of xenobiotics. To date, little is known about monocot UGTs compare to their dicot counterparts, despite their potential role in defence and modification of health-promoting component of cereals essential to human diet. This thesis focuses on identification and functional investigation of UGTs expressed in in the diploid oat species *Avena strigosa*.

Chapters 1 and 2 consist of the General Introduction and Material and Methods, respectively. In chapter 3, a systematic analysis of root-expressed UGTs was carried out using transcriptomic and proteomic approaches. A subset of UGTs was then selected for biochemical analysis. Of particular interest were candidates for glycosylation of avenacin, an antimicrobial triterpenoid glycoside that protects oat against fungi infection.

In chapter 4, the sugar donor specificity of the recombinant UGTs and their activity towards different triterpenoid acceptors was investigated. In chapter 5, a transient expression system was established in *Nicotiana benthamiana* in order to investigate UGT activity. Heterologous co-expression of UGTs with early avenacin biosynthetic pathway enzymes leads to biosynthesis of new-to-nature triterpenoid glycosides, so providing a powerful system for functional analysis of terpene glycosylation *in planta*.

In chapter 6, the catalytic properties of the UGT collection towards different plant natural products was investigated, leading to the production of glycosides of interest. The contribution of this study to the understanding of the evolution and function of monocot UGTs and to their potential commercial exploitation is discussed in the chapter 7.

Acknowledgement

Contents

List of figures

IX

List of tables

Chapter 1 – General introduction

1.1 Plant secondary metabolism

1.1.1 Introduction to plant secondary metabolism

Plants expend considerable energy and assimilated carbon on synthesis of secondary metabolites. In a few cases secondary metabolites have been shown to have functions in plant growth and development, for example flavonoids modulate auxin transport (Peer and Murphy 2007; Sonderby et al. 2010; Vetter 2000). However, the main function of these molecules is likely to be in mediating interactions between plants and the environment (Hartmann 2007; Wink 2003). Metabolic diversification in plants is likely to have occurred because the ability to synthesise specialised secondary metabolites provides selective advantages that enable plants to survive in different ecological niches. These molecules have been variously associated with abiotic stress resistance (Nuccio et al. 1999; Trossat et al. 1998), defence against herbivores and pathogens (Howe and Jander 2008; Osbourn 1996; Vetter 2000; Yang et al. 2013), establishment of symbiotic interactions (Oldroyd 2013), allelopathy (Weston and Mathesius 2013), and attraction of pollinators (Ogata et al. 2005; Theis and Raguso 2005; Yu and Utsumi 2009).

Major classes of plant secondary metabolites include terpenes, phenylpropanoids and alkaloids. The remarkable diversity of plant secondary metabolites arises from a restricted number of carbon skeletons. For example, more than 20,000 reported triterpenoids are derived from 200 triterpene skeletons (Phillips et al. 2006; Xu et al. 2004) . Modification of these core skeletons has led to a unique metabolite profile specific to each plant species. Glycosyltransferases, methyltransferases, cytochrome P450s, acyltransferases and other tailoring enzymes are involved in the decoration of these skeletons (Bak et al. 2011; Bowles et al. 2006; Mugford and Milkowski 2012). The importance of these enzyme families in secondary metabolite diversity is reflected by the marked amplification and diversification of the corresponding gene families within the genomes of higher plants (Caputi et al. 2011; Omura 2013).

Roots accumulate a wide variety of specialised metabolites, including flavonoids, terpenoids and cyanogenic compounds. These chemicals have important roles in defence against soil-borne pathogens, allelopathy, interaction with symbiotic organisms, nutrient uptake and also phytohormones-mediated signalling (D'Auria and Gershenzon 2005; Flores et al. 1999; Weston and Mathesius 2013). The Osbourn laboratory is particularly interested in avenacin, an antimicrobial terpenoid glycoside accumulated in oat root tips.

1.1.2   A large part of secondary metabolites are glycosylated

Glycosides form a large proportion of plant secondary metabolites. Glycosylated flavonoids, triterpenoids (saponins) or cyanogenic compounds are widespread in plant kingdom (Liu et al. 2013; Vetter 2000; Vincken et al. 2007). Glycosylation takes a great part in structural diversification of secondary metabolites. Glycosidic moieties may be composed of a single sugar residue (hexose or pentose) or sugar residues may assemble to form linear or branched sugar chains (Augustin et al. 2011; Heim et al. 2002). The great diversity of glycosylated patterns decorating secondary metabolites may be illustrated by the 300 glycosides reported for the single flavonol, quercetin (Harborne and Baxter 1999). Glycosylation modifies reactivity and solubility of the corresponding aglycones impacting cellular localization and bioactivity. The nature of the glycosides may be storage form of bioactive compounds like benzoxazinoids, glucosinolates or cyanogenic glycosides (Sonderby et al. 2010; Vetter 2000; von Rad et al. 2001); glycosylation may be required for biological activity like membrane permeabilisation induced by saponins (Armah et al. 1999; Augustin et al. 2011) or UV-B protection provided by flavonols glycosides (Liu et al. 2013; Ono et al. 2010a).

1.2   Glycosylation of plant secondary metabolites

1.2.1   Glycosylation is a common feature of plant secondary metabolites

Glycosyltransferases (GTs) are required for the transfer of monosaccharide moieties onto a variety of molecules (e.g. proteins, lipids, and polysaccharides). GTs form a large enzyme superfamily and have considerable diversity in their primary structures (Hu and Walker 2002), they are classified into 94 different enzyme families (http://www.cazy.org/GlycosylTransferases.html) (Campbell et al. 1997; Cantarel et al. 2009). The majority of these enzymes use nucleoside diphosphate activated sugars as donors and are therefore uridine diphosphate-dependant glycosyltransferases (UGTs). Glycosyltransferases belonging to family 1 (GT1s) are UGTs involved in the glycosylation of natural products (Bowles et al. 2006; Vogt and Jones 2000).

UGTs make major contributions to the diversity of plant natural products (Bowles et al. 2005; Bowles et al. 2006; Gachon et al. 2005b; Vogt and Jones 2000; Yonekura-Sakakibara and Hanada 2011). They have a variety of functions. For example, UGTs are considered to be important in regulating plant growth and development due to their activity towards phytohormones (Husar et al. 2011; Jin et al. 2013; Lim et al. 2005; Martin et al. 2001; Martin et al. 1999; Priest et al. 2006; Tognetti et al. 2010; Wang et al. 2013a), conjugated phytohormones are regarded as storage, transport or inactivated forms of their unconjugated counterparts and are essential to hormonal homeostasis (Bajguz and Piotrowska 2009; Piotrowska and Bajguz 2011). UGTs are also important players in detoxification metabolism (Brazier-Hicks et al. 2007a; Brazier-Hicks et al. 2007b; Meech et al. 2012; Messner et al. 2003). They are responsible for glucoconjugation of toxic compounds preceding their transfer and storage in the vacuole, rendering them inert; alternatively xenobiotic glycosides may also be exported outside of the cell (Brazier-Hicks et al. 2007b; Cole and Edwards 2000). UGTs also participate to the elaboration of bioactive molecules important in defence mechanisms like terpenoid glycosides, glucosinolate, cyanogenic glycosides, or flavonoids glycposides (Agati et al. 2011; Grubb et al. 2004; Kannangara et al. 2011; Klee 2013; Sawai and Saito 2011). Modification of solubility, stability and volatility through glycosylation will affect sequestration of plant compounds within the cell. The stability of pigments

(e.g. betalaines, anthocyanins), modification of the taste of fruit flesh (e.g. naringenin, saffron, steviosides) and retention of aromas in flowers or fruits (e.g. monoterpenes, phenyl alcohols) are all examples of the ways in which modification of the properties of molecules by UGTs can influence plant interactions with the environment (Frydman et al. 2013; Madhav et al. 2013; Moraga et al. 2009; Moraga et al. 2004; Nishihara and Nakatsuka 2011; Ono et al. 2006; Ono et al. 2010b; Sui et al. 2011).

UGTs catalyse the transfer of a sugar from a sugar donor (a uridine diphospho (UDP)-sugar) to a sugar acceptor (usually a lipophilic molecule) (Fig. 1.1). GT1 are inverting glycosyltransferases, those enzymes will invert the stereochemistry of the anomeric carbon at position C1 of the sugar residue (Lairson et al. 2008; Lairson and Withers 2004). UGTs generally show high specificity for their sugar donors and a single activated sugar (UDP-sugars) is efficiently recognised as the substrate (Bowles et al. 2006; Kubo et al. 2004; Noguchi et al. 2009; Osmani et al. 2008). UGTs can be either very selective or promiscuous in term of the range of acceptors they recognise. The major principle governing acceptor recognition by UGTs seems to be regioselectivity (systematic glycosylation of the same position) rather than specificity over one or few structurally related compounds (Cartwright et al. 2008; Hansen et al. 2003; Vogt and Jones 2000). Most UGTs act on hydroxyl or carboxyl groups but *N*-glycosylation, *S*-glycosylation or *C*-glycosylation can also occur (Brazier-Hicks et al. 2009; Brazier-Hicks et al. 2007b; Grubb et al. 2004; Hou et al. 2004; Jones and Vogt 2001; Wang et al. 2011).
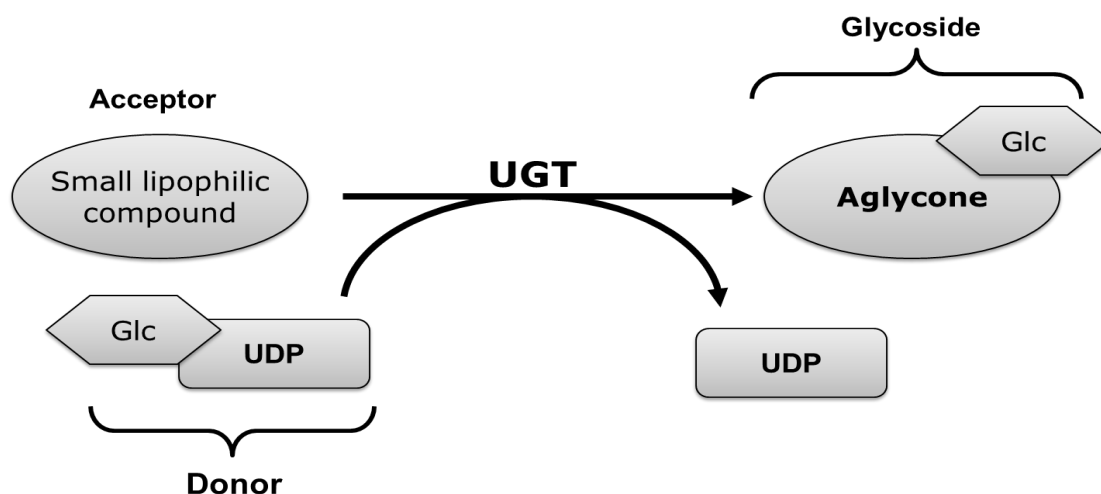
Figure 1.1: Simplified diagram of the glycosylation reaction catalysed by UGTs. Family 1 glycosyltransferases (UGTs) catalyse the transfer of a sugar (shown here as glucose; Glc) from an activated sugar donor (uridine diphosphate glucose; UDP) onto a lipophilic acceptor to form a glycoside. UDP is released during the process.

1.2.2 Activity of plant UGTs

Recent release of crystallographic data from several plant UGTs has enhanced our understanding of glycosylation machanism and enabled homology-based 3D modelling of a number of UGTs (Osmani et al. 2008; Osmani et al. 2009; Wang 2009). UGT71G1 of *Medicago truncatula* was the first plant UGT structure to be solved (Shao et al. 2005). Four other crystal structures have subsequently been published (Brazier-Hicks et al. 2007b; Li et al. 2007; Modolo et al. 2009; Offen et al. 2006). Despite the fact that those enzymes have relatively low amino acid sequence identity (e.g. 25-45%) they share very similar 3D structures (Osmani et al. 2009; Wang 2009).

These studies have clarified the overall architecture of plant UGTs as a GT-B fold. GT-B is one of the two structural folds identified for sugar nucleotide dependent enzymes. The GT-B structure was first described in 1994 with the resolution of the structure of the bacteriophage T4 β-glucosyltransferase (Vrielink et al. 1994). The GT-B fold is formed of two Rossmann-like domains, each composed of central β-strands surrounding by several α-helices. The catalytic site is localized in a cleft between the two domains (Fig. 1.2.B). In plant UGTs, two highly conserved residues play a crucial part in the S_N2-like mechanism similar to other inverting GTs (Lairson et al. 2008; Osmani et al. 2009; Wang 2009). A histidine positioned around the twentieth residue acts as a general base to deprotonate the acceptor (Fig. 1.2.C, His22). An aspartate residue linked by a hydrogen bond to the histidine allows the stabilisation of the entire complex after the deprotonation (Fig. 1.2.C, Asp121). Nucleophilic attack of the C1 carbon of the UDP-sugar is achieved by the deprotonated acceptor (Brazier-Hicks et al. 2007b; Lairson et al. 2008; Shao et al. 2005; Wang 2009). The sugar donor binding site is formed by a highly conserved motif throughout the UGT family and is localized in the C-terminal domain of the enzyme (Fig 1.2.A). The acceptor binding site is opposite to the donor binding site in the N-terminal part of the protein (Shao et al. 2005; Wang 2009).
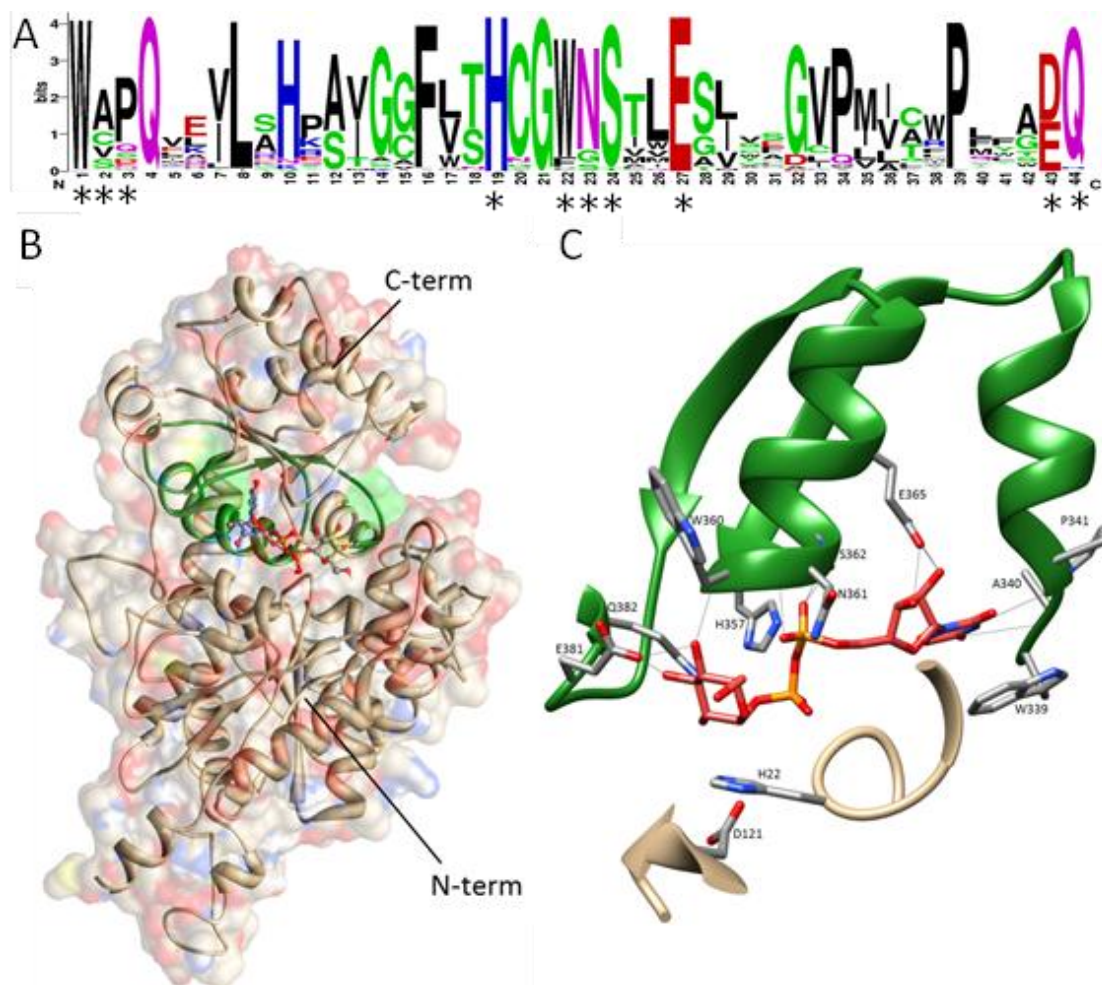
Figure 1.2: Structural basis of sugar recognition by UGTs. A. Consensus PSPG motif of *A. thaliana*. Sequences were retrieved from http://www.p450.kvl.dk/UGT.shtml. A graphical representation of residues constituting the PSPG motifs was made using http://weblogo.berkeley.edu/ software. The overall height the stack indicates the sequence conservation at that position, while the height of the symbols within the stack indicates the relative frequency of each amino acid at that position. Essential residues for UDP-sugar binding are labelled with an asterisk. **B.** SAD10 homology model generated using I-Tasser (Zhang et al. 2008). **C.** A view of the 3D structure of the PSPG motif of UGT71G1crystallised in complex with UDP-Glc. The PSPG motif is represented as a green ribbon and the essential residues are labeled as well as the two catalytic residues His22 and Asp121. The chemical structure of UDP-Glc appears in red. Hypothetical hydrogen bonds are shown as grey lines.

1.2.3   Recognition of the sugar donor

Plant UGTs recognise their sugar donors via a motif localized on the C-terminal part of the enzyme (Fig. 1.2.A). This Plant Secondary Product Glycosyltransferase (PSPG) motif is highly conserved throughout UGT families (Hughes and Hughes 1994; Mackenzie et al. 1997; Ross et al. 2001). UDP-glucose (UDP-Glc) is the most common donor, but UDP-galactose (UDP-Gal), UDP-rhamnose (UDP-Rha), UDP-xylose (UDP-Xyl), UDP-glucuronate (UDP-GlcA), UDP-mannose (UDP-Man) or UDP-arabinose (UDP-Ara) are also recognised by plant UGTs (Bowles et al. 2006; Osmani et al. 2009).

UGTs show a high specificity towards their sugar donor mainly due to slight differences in PSPG box sequences (Offen et al. 2006). Determination of donor specificity via the PSPG sequence has been studied and some essential amino acids that are required for donor specificity are now known (Kubo et al. 2004; Noguchi et al. 2009; Osmani et al. 2008). The essential residues of the donor binding pocket have been identified by crystallography and mutational analysis. Residues interacting with UDP are part of the highly conserved portion of the PSPG motif (Fig. 1.2.A and C). The first tryptophan of the PSPG motif (residues of PSPG motif numbered in figure 1.2.A) forms a hydrophobic platform stacking with the uracil ring of UDP (Fig. 1.2.C, W339). The fourth glutamine of the motif and the twenty-seventh glutamate (Fig. 1.2.C, E365) form hydrogen bonds with ribose hydroxyl groups. The nineteenth histidine and twenty-fourth serine interact with the oxygens of the two phosphates (Fig. 1.2.C, H357 and S362). The two last residues of the PSPG motif (Fig. 1.2.C, E381 and Q382) are implicated in sugar recognition by interacting with the hydroxyl groups on position C2, C3 and C4 of glucose (Shao et al. 2005; Wang 2009). The presence of glutamine or histidine as a last residue has been shown to be essential to determine specificity toward UDP-Glc or UDP-Gal (Kubo et al. 2004). More recently, engineering of the flavonoid-7-*O*-glucuronosyltransferase UGT88D7 by introducing only two point mutations changed the preferred sugar-donor from UDP-GlcA to UDP-Glc (Noguchi et al. 2009). These two mutations affected the twenty second tryptophan of the PSPG motif, which interacts with the 4-OH of Glc, and the Ser127. Residues outside of the PSPG motif are also involved in sugar donor binding (Osmani et al. 2009). The Arg25 in the N-terminus of UGT94B1 has been shown to determine the specificity between UDP-GlcA and UDP-Glc (Osmani et al. 2008). Point mutation of this residue to Ser leads to a dramatic decrease of activity with UDP-GlcA and a small increase of the activity towards UDP-Glc. Several UGT rhamnosyltransferases have now been identified in various plant species, sequences analysis suggests a convergent evolution of rhamnosyltransferase activity across plant species and UGT groups. *A. thaliana* rhamnosyltransferases UGT89C1 and UGT78D1 do not possess the conserved glutamine at the end of the PSPG motif which is replaced by a histidine or an asparagine respectively (Jones et al. 2003; Yonekura-Sakakibara et al. 2007). The same asparagine is present in potato StGT3, but is absent in all of the four other rhamnosyltransferases bearing the conserved glutamine residue (Frydman et al. 2013; Itkin et al. 2011; Shibuya et al. 2010).

1.2.4 Acceptor recognition - Towards understanding regiospecificity of glycosylation by UGTs

The mechanism of sugar acceptor recognition by UGTs is poorly understood (Osmani et al. 2009; Wang 2009). *In vitro* studies using recombinant enzymes from *Arabidopsis thaliana* have highlighted the promiscuity of many UGTs with regard to acceptor recognition (Caputi et al. 2008; Lim et al. 2002; Vogt and Jones 2000). Recent work on UGT85K4 and UGT85K5 from cassava suggests a role for these UGTs in cyanogenic glucoside biosynthesis. The linamarin and lotaustralin glycosylation activities of recombinant UGT85K4 and UGT85K5 as assessed *in vitro* are consistent with their role *in planta*; nevertheless, UGT85K4-5 recognises a wide range of other acceptors including flavonoids, isoflavonoids, simple alcohols and various hydroxynitriles (Kannangara et al. 2011). In contrast, some UGTs possess high specificity towards one or a few structurally related compounds. UGT85A24 from *Gardenia jasminoides* proved to be highly specific towards 7-deoxyloganetin but did not recognise an unmethylated form of this compound (Nagatoshi et al. 2011). Modalities for acceptor recognition and binding onto the N-terminal part of the protein remain elusive. Acceptor binding pockets of plant UGTs are generally surrounded by apolar residues (Modolo et al. 2009; Wang 2009). Residues within these acceptor pockets are located in regions that are poorly conserved through plant UGTs (Osmani et al. 2009). Little is known about the mechanisms underlying the correct orientation of the acceptor molecule and therefore the regiospecificity of glycosylation catalysed by UGTs. Steric constraints due to the nature of the surrounding amino acids and formation of H-bonds are believed to be involved in acceptor binding (He et al. 2006; Li et al. 2007; Modolo et al. 2009; van der Heide 1966).

1.2.5   Phylogeny of plant UGTs, evolution and function prediction

The family 1 GTs are one of the largest groups of natural product-decorating enzymes in higher plants. The expansion of this family in higher plants reflects chemical diversification during the adaptation of plants to life on land (Caputi et al. 2011; Yonekura-Sakakibara and Hanada 2011). UGTs sharing 40% or more amino acid identity have been classified into families, the plant UGT families extending from UGT71 to UGT100 (Mackenzie et al. 1997). Mining of the complete genome sequence of thale cress (*Arabidopsis thaliana*) has identified 107 predicted family 1 GT genes. These GTs have been classified into 14 monophyletic groups (groups A to N; Fig. 1.3) (Ross et al. 2001). Recently two new phylogenetic groups (O and P) which are not present in *A. thaliana* but are represented in other plant species have also been reported (Caputi et al. 2011).

Specificity for each sugar donor appears to have evolved multiple times and closely related enzymes may use different sugar donors. This is illustrated by the three *A. thaliana* flavonoid glycosyltransferase homologues UGT78D1, UGT78D2 and UGT78D3, each of which use a different UDP-sugar as a donor. UGT78D1 has flavonoid-3-*O*-rhamnosyltransferase activity *in planta* (Jones et al. 2003), UGT78D2 is a flavonoid-3-*O*-glucosytransferase (Yonekura-Sakakibara et al. 2007) and UGT78D3 is a flavonoid-3-*O*-arabinosyltransferase (Yonekura-Sakakibara et al. 2008).

UGTs acting on similar acceptors appear to be spread across the phylogenetic tree. UGTs belonging to the same monophylogenic group normally share the same regiospecificity for glycosylation of their acceptors (Lim et al. 2004; Vogt and Jones 2000). A large dataset has been collected for flavonoid glycosyltransferases; these GTs are predominantly part of 4 discrete phylogenetic groups matching with the regiospecificity of these enzymes (Frydman et al. 2013; Noguchi et al. 2009; Yonekura-Sakakibara et al. 2012). Far less is known about triterpenoid glycosyltransferases.
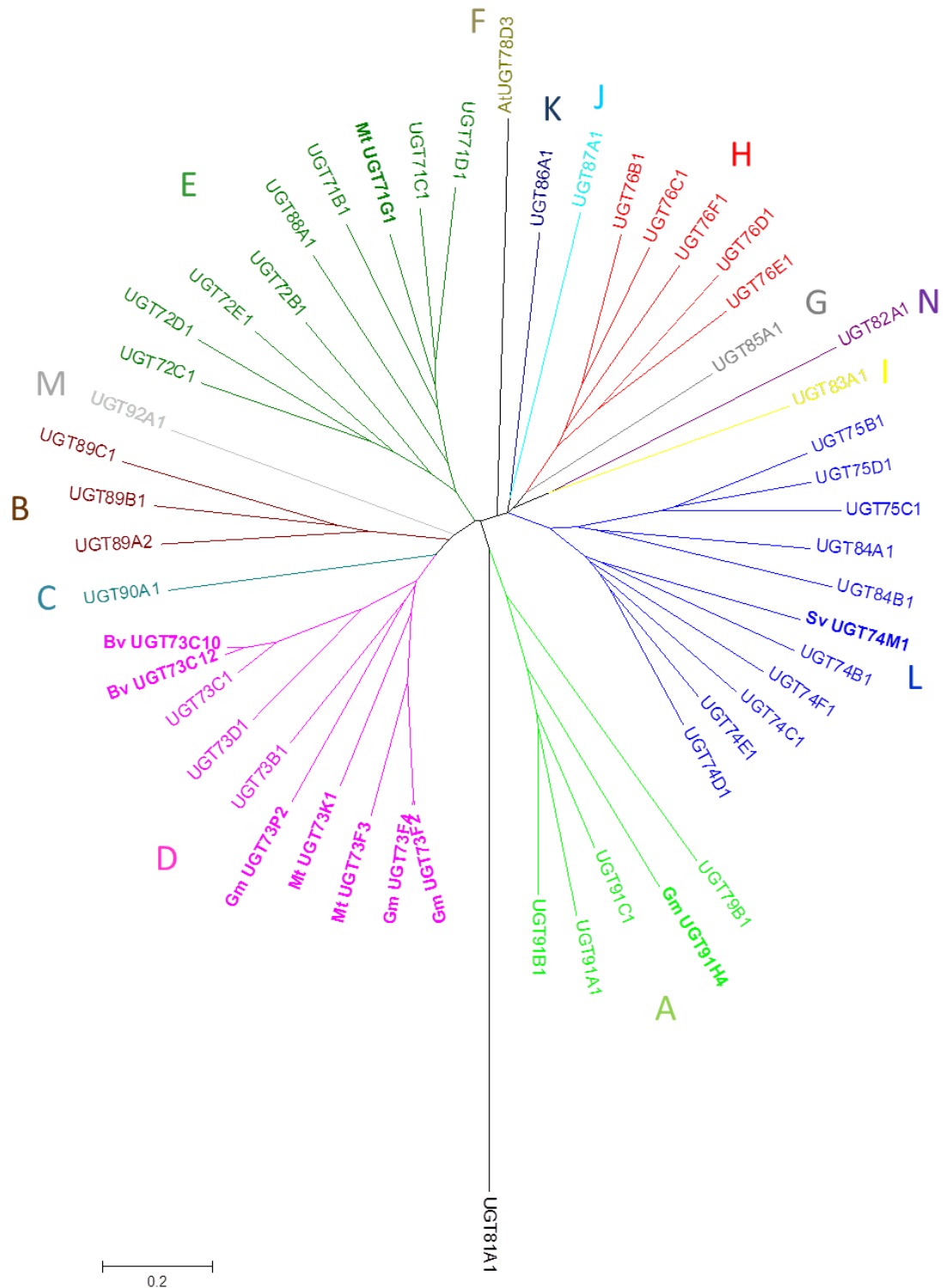
Figure 1.3: Phylogenetic tree of amino acid sequences of *A. thaliana* UGTs, also including characterised triterpenoid glycosyltransferases from other plant species. The monophyletic groups defined by Li et al. (2001) are labelled on the tree. Triterpenoid glycosyltransferases are indicated in bold. The phylogenetic tree was drawn using the Neighbor-Joining method. Selected family one glycosyltransferases from *A. thaliana* were retrieved from http://www.p450.kvl.dk/UGT.shtml. The accession numbers of the triterpenoid glycosyltransferases from other plant species are given in supplementary data (S.3). The scale bar represents 20% divergence.

1.3 Terpenoids and triterpenoid glycosides

1.3.1 The plant terpenome

Terpenoids are one of the largest groups of metabolites found in nature. More than 50,000 terpenoids have been reported, most of them in plants, fungi or bacteria (Cane and Ikeda 2012; Hemmerlin et al. 2012). Some plant terpenoids have essential physiological functions, for example in cell membrane structure and fluidity (sterols), hormone signalling (gibberellins, brassinosteroids) and photosynthesis (carotenoids) (Daviere and Achard 2013; Simons and Sampaio 2011; Sozer et al. 2011; Zhu et al. 2013). However the majority of higher plant terpenes are likely to have ecological functions, including plant defence (diterpenoids, sesquiterpenoids, triterpene glycoside saponins), scents and aromas used as attractants (monoterpenoids, sesquiterpenoids), or allelopathy (diterpenoids, triterpenoids) (Augustin et al. 2011; Dixon 2001; Gonzalez-Lamothe et al. 2009; Hemmerlin et al. 2012; Yu and Utsumi 2009).

The vast structural diversity found in terpenoids originates from a common precursor isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP). These initial terpene building blocks can both be synthesised via two independent pathways, the mevalonate (MVA) pathway and the 2C-methyl-D-erythritol 4-phosphate (MEP) pathway (Hemmerlin et al. 2012) (Fig. 1.4). The MVA pathway takes place in the cytosol and uses acetyl Co-A as a precursor. The MEP pathway by contrast is plastidial and uses pyruvate and D-glyceraldehyde 3-phosphate as precursors, which originate indirectly from the Calvin cycle. Isoprenyl diphosphate synthases catalyse condensation of IPP and DMAPP to form the linear isoprene backbones of each terpene family (Wang and Ohnuma 2000). Isoprene diphosphate precursors are then processed by terpene synthases (Chen et al. 2011) to form mono- (C10), sesqui- (C15), di-(C20), tri- (C30) and tetraterpenoids (carotenoids) (C40). In the plastid, IPP and DMAPP derived from the MEP pathway are used to form geranyl diphosphate (GPP, C10), the precursor of monoterpenes, and geranygeranyl diphosphate (GGPP, C20), the precursor of diterpenes. Condensation of two GGPP molecules leads to formation of phytoene, the precursor of the carotenoids. In the cytosol, the MVA pathway generates farnesyl diphosphate (FPP, C15) via FPP synthase. FPP is the direct precursor of sesquiterpenes.

Alternatively, condensation of two FPP molecules and epoxidation of the resulting C30 molecule squalene leads to formation of 2-3-oxidosqualene, the precursor of sterols and triterpenes (Chung et al. 2013; Hemmerlin et al. 2012; Misawa 2011). Sterols are part of the primary metabolism and their three first rings A, B and C are in chair-boat-chair conformation. Triterpenes are considered as secondary metabolites and have a chair-chair-chair conformation (Thimmappa et al. 2014).
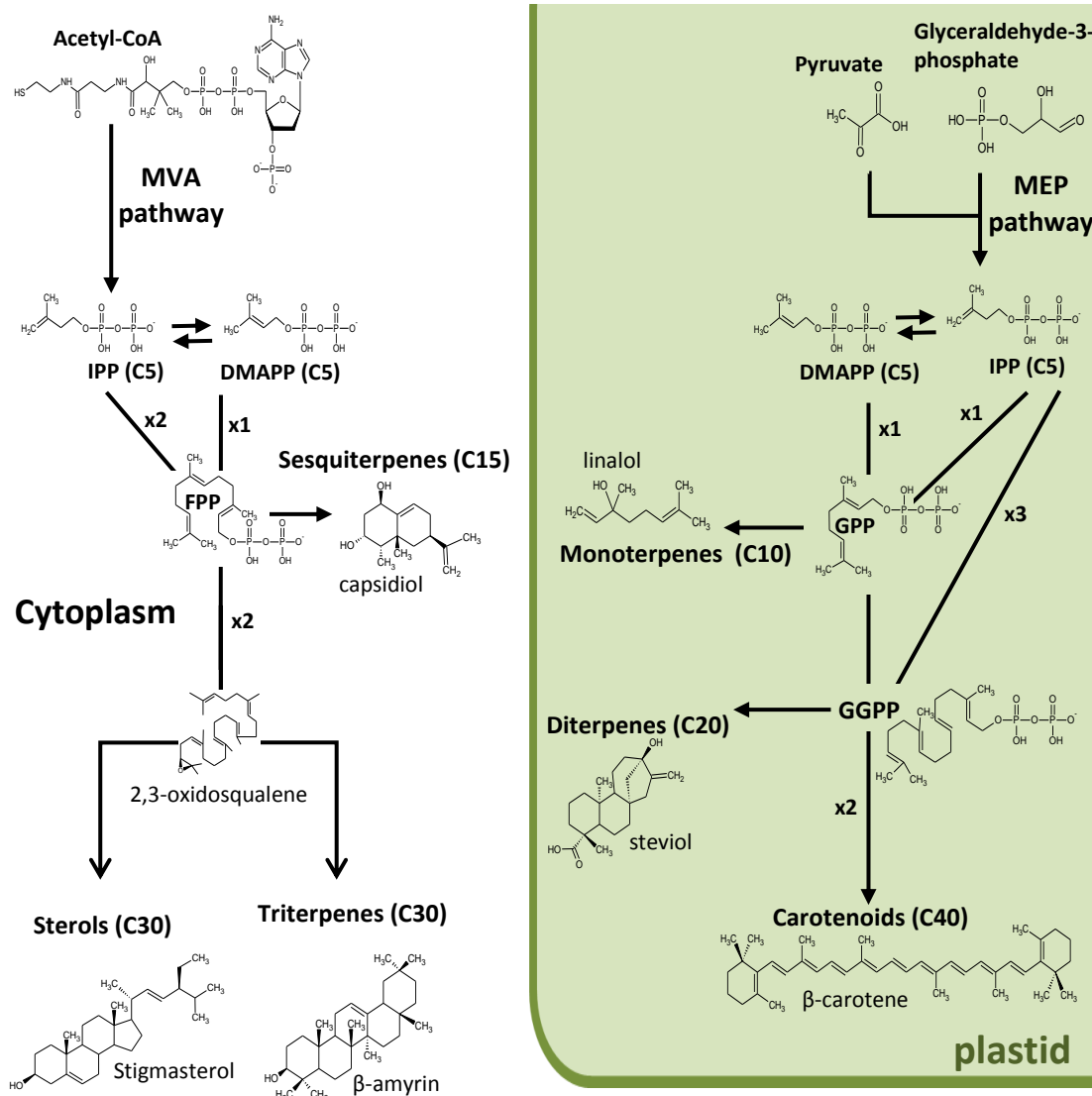
Figure 1.4: Terpene biosynthesis in plants. The cytosolic mevalonate (MVA) pathway and plastidial 2-C-methylerythritol 4-phosphate (MEP) pathway co-exist in plant cells. Both pathways form dimethylallyl diphosphate (DMAPP) and isoprenyl diphosphate (IPP), the precursors of all terpenes. Condensation of these precursors leads to the formation of linear backbones for all terpenoid compounds; geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP). Further condensation is required to synthesised oxidosqualene (2x FPP) and phytoene (2x GGPP), the respective precursors of triterpenes/sterols and carotenoids. Terpene synthases use the linear precursors GPP, FPP and GGPP to form monoterpenes, sesquiterpenes and diterpenes, respectively.

1.3.2   Glycosylated triterpenes (saponins)

Saponins are secondary metabolites that are found in a wide range of plant species. They are glycosylated triterpenes or sterols that have surfactant properties due to their amphiphilic nature. They are likely to have functions in protection of plants against attack by pathogenic microbes, herbivores and competing plant species (Augustin et al. 2011; Osbourn 1996; Sawai and Saito 2011).

The defensive role of saponins has been attributed to their ability to permeabilize plasma membranes. The amphipathic properties of these molecules allow them to penetrate the external monolayer of the cellular membrane. After integration into the membrane they then associate with sterols and form complexes. Complex formation results in deformation of the membrane and ultimately to membrane disruption through pore formation or vesiculation (Fig. 1.5), and so to the lysis of the cell (Armah et al. 1999; Augustin et al. 2011; Keukens et al. 1995; Nishikawa et al. 1984).

Saponins have considerable commercial importance (table 1.1). Historically, they were used as soaps due to the surfactant properties of plant extracts with high saponin content. For example, leaf extracts of *Saponaria officinalis* were used a long time ago as detergent to wash linen (Osbourn 1996) and saponins from the bark of *Quillaja saponaria* were used as shampoo by South American tribes (Francis et al. 2002). Saponins are also the major active components of many traditional medicines, including roots preparations from *Panax* spp (ginseng), *Panax notoginseng*, *Symplocos chinensis* … (Sparg et al. 2004; Yang et al. 2009). Many saponins have positive effects on human health, including anti-tumoral or anti-inflammatory activities (Dinda et al. 2010; Sparg et al. 2004; Vincken et al. 2007). Some saponins are also potent immunoadjuvant compounds, in particular QS-21 from *Quillaja saponaria* is a widely used saponin-based adjuvant (Sun et al. 2009). Saponins are also exploited by the cosmetic and food industries, and steroidal saponins from *Yucca schidigera* are used extensively as anti-food-deteriorating agents (Sparg et al. 2004; Tanaka et al. 1996).
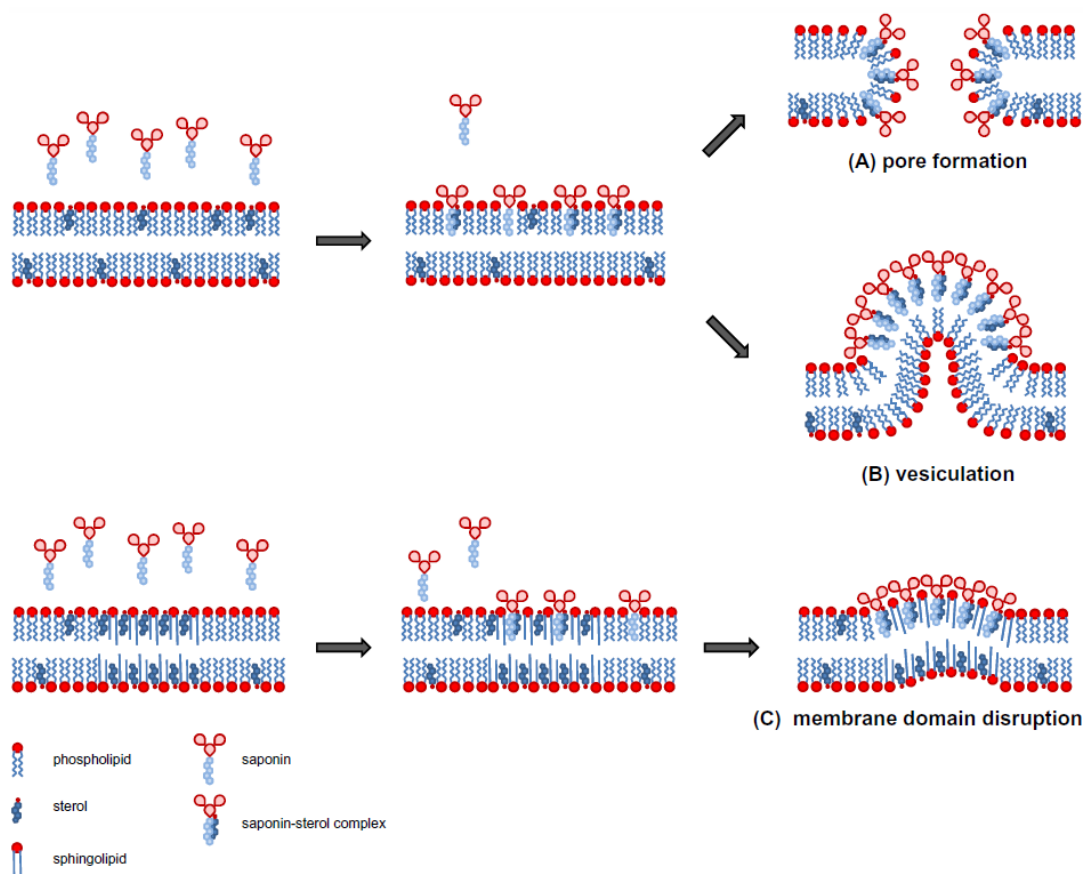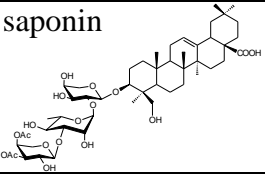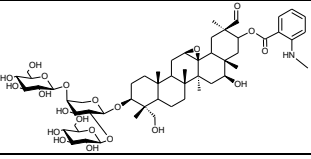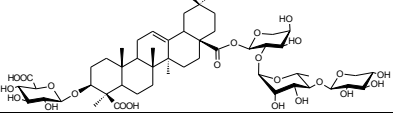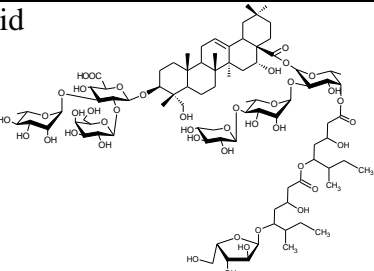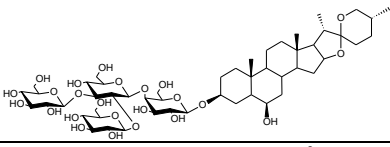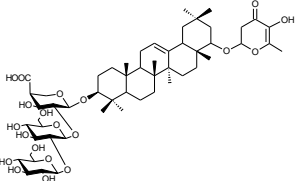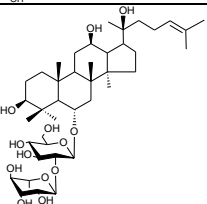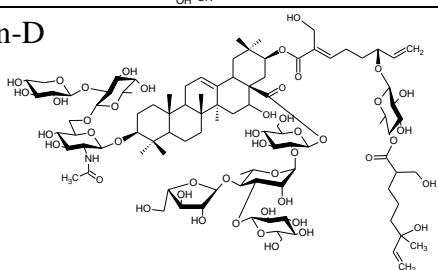
Figure 1.5: Schematic models of the molecular mechanisms of saponin activities towards membranes. Saponins integrate with their hydrophobic part (sapogenin) into the membrane. Within the membrane they form complexes with sterols, which subsequently, driven by interaction of their extra-membranous orientated saccharide residues, accumulate into plaques. Steric interference of these saccharide moieties causes membrane curvature subsequently leading to: (A) pore formation in the membrane, or (B) hemitubular protuberances resulting in sterol extraction via vesiculation. Alternatively, after membrane integration saponins may migrate towards sphingolipid/sterol enriched membrane domains (C) prior to complex formation with the incorporated sterols, thereby interfering with specific domain functionalities. Figure copied from Augustin et al. 2011.

Triterpenes and steroidal saponins are derived from the MVA pathway and share a common precursor, 2,3-oxidosqualene. Cyclisation of 2,3-oxidosqualene is the first step in saponin biosynthesis, leading to formation of the carbon skeleton of the molecule. This core structure is then decorated by different tailoring enzymes. Cytochrome P450 (CYP450s) are involved in oxidation steps while glycosyltransferases (UGTs) catalyse the formation of sugar chains. These two classes of enzyme form a major part of saponin biosynthetic pathways but other tailoring enzymes (e.g. methyltransferases and acyltransferases) may also be involved (Augustin et al. 2011; Mugford et al. 2013; Sawai and Saito 2011).

Table 1.1: Biological function and human use of plant saponins.

Plant species	Saponin	Biological function, human use
Sapindus mukorossi	*S.mukorossi* saponin	Molluscicidal activity Detergent
Avena strigosa	Avenacin A-1	Antifungal activity
Medicago truncatula	3-GlcA-28-Ara-Rha-Xyl-medicagenate	Insecticidal activity
Quillaja saponaria	Quillaic acid	Digestion improved in ruminants
Allium sativa	Eruboside-B	Protection against cardiovascular disease
Glycine max	Soyasaponin βg	Antitumor activity
Panax ginseng	Ginsenoside-Rg2	Protection against memory impairment
Acacia victoriae	Avicin-D	Antioxidant activity

References from: Sparg et al. 2004, Vincken et al. 2007, Guclu-Ustundag and Mazza 2007.

Cyclisation of 2,3-oxidosqualene is the branch-point between the sterol and triterpene pathways. This step is catalysed by oxidosqualene cyclases (OSCs). Most OSCs give rise to polycyclic triterpenes by cyclisation cascades involving various cationic intermediates (Xu et al. 2004). The initial cation formed by triterpene-synthesising OSCs is the dammarenyl cation giving rise to the common chair-chair-chair conformation of triterpenes. In contrast, the protosteryl cation of cycloartenol and lanosterol synthases gives rise to the chair-boat-chair conformation of sterols (Xue et al. 2012). Oleanane, also referred to as β-amyrin, is the most common plant triterpene skeleton, but saponins derived from other triterpene skeletons (e.g. lupeol and α-amyrin) are also found (Connolly and Hill 2010; Vincken et al. 2007).

Various oxidation steps occurring during saponin biosynthesis are catalysed by CYP450s (Augustin et al. 2011; Sawai and Saito 2011). CYP450s are one of the most expansive gene families in higher plants (Bak et al. 2011; Nelson 2013; Omura 2013). These enzymes are monooxygenases. Their catalytic centre is composed of heme, with iron involved in binding an oxygen molecule. Following activation via electron transfer (catalysed by NADPH-cytochrome P450 reductases), canonical CYP450s catalyse the oxidation of hydrocarbon molecules (Bak et al. 2011). In the case of saponin biosynthesis, nineteen P450s from three distinct clans (CYP71 CYP72 and CYP85; clans are defined according to the percentage of identity between CYP450s, ≥40% in same clan) have been functionally characterised in dicot species. They catalyse the formation of hydroxyl, carbonyl and epoxy functional groups of the triterpene backbone of saponins (Kunii et al. 2012; Seki et al. 2008; Shibuya et al. 2006). Recently the first monocot triterpene P450 (CYP51H10) was characterised in our lab. This belongs to the CYP51 clan (Geisler et al. 2013). CYP51H10 is able to catalyse hydroxylation and epoxidation of β-amyrin. Oxidations catalysed by CYP450s will change the solubility of the triterpene skeleton, increasing its accessibility for other tailoring enzymes and providing functional groups (hydroxyl, carboxyl) likely to be targeted by glycosyltransferases.

Glycosylation is an essential characteristic of all saponins. The addition of a glycosidic moiety to the triterpenoid aglycone substantially modifies the properties of the molecule by adding a polar group onto the apolar backbone of the aglycone (sapogenin). The amphipathic nature of the resulting product is crucial for bioactivity (Armah et al. 1999; Augustin et al. 2011). The formation of oligosaccharide chains in the synthesis of triterpene glycosides and saponins in general is likely to occur by sequential addition of one monosaccharide residue at a time catalysed by family one glycosyltransferases (Augustin et al. 2011; Sawai and Saito 2011). The large action in the glycosylation of sapogenins is a source of considerable diversity in saponin structures. The number of sugar chains, their composition and their position on the triterpenoid backbone provides many possible combinations for a single aglycone (Vincken et al. 2007).

Triterpenoid saponins sugar chains are usually composed of D-glucose, D-galactose, L-arabinose, L-rhamnose, D-xylose, D-mannose and D-glucuronic acid; other rare monosaccharide residues such as D-apiose, L-fucose, ribose or D-quinovose have also been found in some plant species (Vincken et al. 2007). These residues may be attached alone to the aglycone or they can form part of a sugar chain. The sugar chains of triterpenoid saponins can be composed of up to 8 residues but the norm is between 3 and 5 residues. These sugar chains may be linear or branched (table 1.1). Most of the monosaccharides are linked to carbon 3 or carbon 28 of the triterpenoid backbone, although other sites of glycosylation exist (C4, C16, C20, C21, C22, C23). Common patterns of glycosylation involve one or two sugar chains (giving monodesmosidic or bidesmosidic saponins, respectively). Rare examples of tridesmosidic saponins (possessing three sugar chains) have also been reported (Bedir et al. 1998; Yesilada et al. 2005). These sugar chains may be added onto hydroxyl groups or carboxyl groups and therefore form sugar acetals or sugar esters, respectively (Vincken et al. 2007).

1.3.3 Glycosylation of triterpenes by UGTs

Despite the central role of glycosylation in structural diversification of saponins and the growing body of knowledge about the structural and functional properties of UGTs, only a handful of triterpenoid glycosyltransferases have been identified so far (table 1.2; Fig. 1.6).

In *Medicago truncatula,* three triterpenoid glycosyltransferases have been identified based on their co-expression with the *M. truncatula β*-amyrin synthase gene in elicited cell cultures. Biochemical analysis of UGT73K1 and UGT71G1, revealed activity towards triterpenoid acceptors (Achnine et al. 2005). A third enzyme, UGT73F3, was shown to have hederagenin-28-*O*-glucosyltransferase activity (Fig. 1.6). *M. truncatula* insertion lines for the gene encoding this UGT had modified saponin content, confirming the role of this UGT in triterpene biosynthesis *in planta* (Naoumkina et al. 2010).

The sequence encoding the *Saponaria vaccaria* UGT, Ugt74M1, was isolated from an expressed sequence tag (EST) collection generated from saponin-producing tissues. Recombinant UGT74M1 is an ester-forming glucosyltransferase of gypsogenic acid that carries out glucosylation at position C-28 (Fig. 1.6) (Meesapyodsuk et al. 2007).

Investigation of sequences that are homologous to known triterpenoid glycosyltransferases from *M. truncatula* have been identified in an EST database for soybean (*Glycine max*) and two glycosyltransferases involved in the biosynthesis of soyasaponin I in *G. max* have been identified (UGT73P2 and UGT91H4; Fig. 1.6). The two recombinant UGTs are involved in the sequential addition of D-galactose then L-rhamnose in the synthesis of soyasaponin I C-3 linked sugar chains (Shibuya et al. 2010). Genetic analysis of a naturally occurring *G. max* accession that produced soyasaponin A lacking terminal sugar at position C-22 has led to identification of a locus that regulates the composition of the saponin sugar chain. This locus corresponds to a *Ugt* gene responsible for the addition of the terminal monosaccharide residue present on the C-22 sugar chain of soyasaponin A (Fig. 1.6). Two alleles of this gene, *Ugt73f2* and *Ugt73f4,*are responsible for addition of D-glucose and L-rhamnose, respectively (Fig. 1.6) (Sayama et al. 2012).

Augustin et al. (2012) reported the identification of four triterpenoid glycosyltransferases UGT73Cs from *Barbarea vulgaris*. An initial screening of a cDNA expression library led to the discovery of a UGT catalysing glucosylation of a mixture of *B. vulgaris* sapogenins. Homologues of this enzyme were found in 454 pyrosequencing-generated transcriptomic database of *B. vulgaris*. UGT73C10, UGT73C11, UGT73C12 and UGT73C13 recombinant enzymes all catalysed 3-*O*-glucosylation of oleanolic acid, hederagenin and betulinic acid (Fig. 1.6). In addition to monoglucosylated products, UGT73C12 and UGT73C13 also formed low amounts of bidesmosidic products.
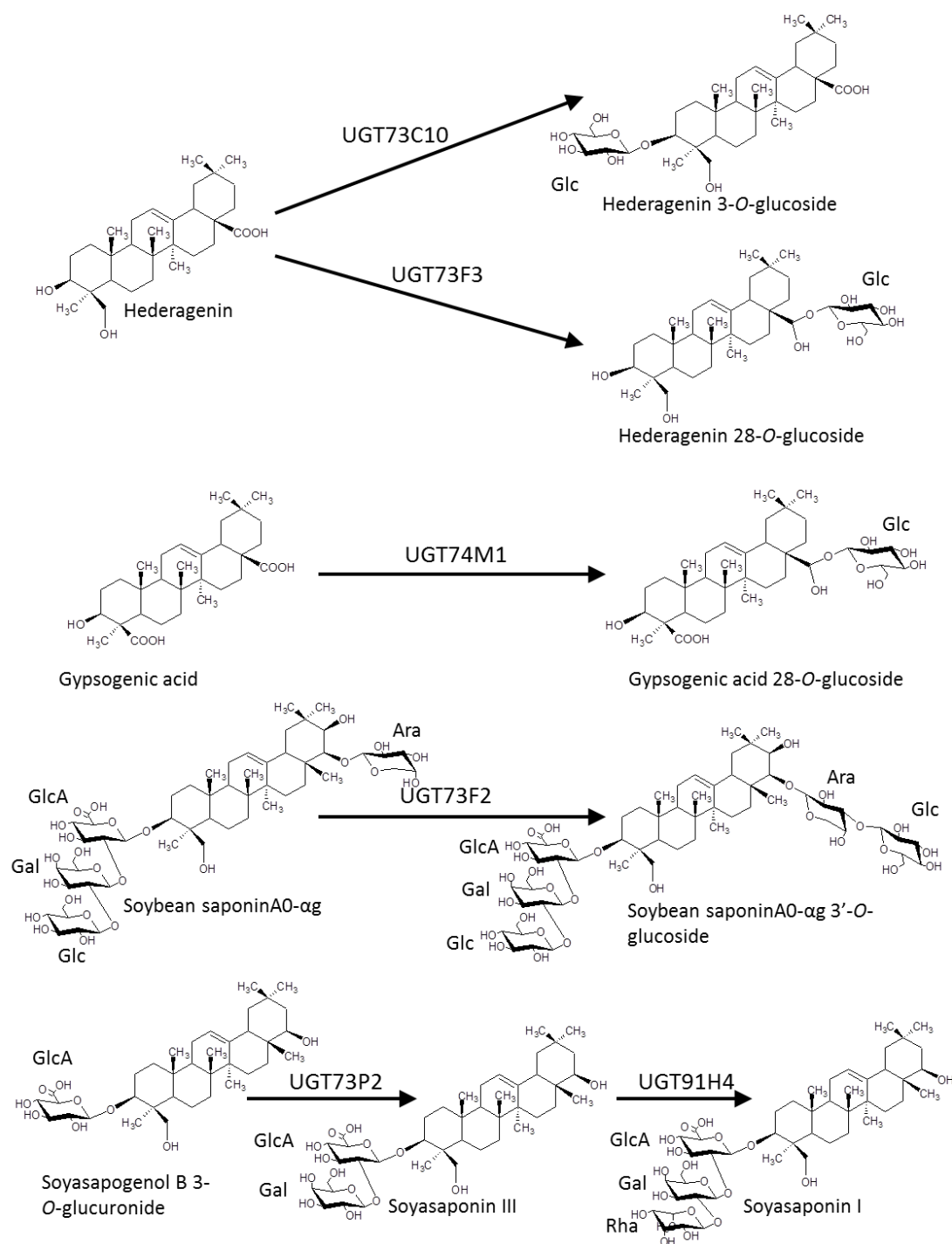
Figure 1.6. Reactions catalysed by characterized triterpene glycosyltransferases. UGT73C10 from *Barbarea vulgaris* and UGT73F3 *from Medicago truncatula* glucosylate the C-3 and C-28 positions of β-amyrin-derived (oleanane) triterpenes, respectively (Augustin et al, 2012; Naoumkina et al, 2010). The Saponaria vaccaria enzyme UGT74M1 glucosylates the C-28 position of another oleanane triterpene, gypsogenic acid (Meesapyodsuk et al, 2007). Three glycosytransferases that add sugars to triterpene glycosides are also shown. UGT73F2 from soybean (*Glycine max*) glucosylates the C-22-linked arabinose of soybean saponin AO-αg (Sayama et al, 2012); UGT73P2, also from soybean, adds a galactose to the C-3 linked glucuronic acid of soyasapogenol-B, and a second soybean enzyme UGT91H4 then adds a rhamnose to the galactose moiety (Shibuya et al, 2010).

Table 1.2. Triterpene glycosyltransferases

Family	Gene Name	Species	GenBank ID	Substrate	Reaction	Reference
71	UGT71G1	*Medicago trunctula*	AAW56092	Hederagenin	β-D-glucosylation	He et al. 2006 ; Shao et al, 2005 ; Achnine et al. 2005
73	UGT73C10	*Barbarea vulgaris*	AFN26666	β-Amyrin, hederagenin	C3- β-D-glucosylation	Augustin et al. 2012
	UGT73C11	*Barbarea vulgaris*	AFN26667	β-Amyrin, hederagenin	C3- β-D-glucosylation	Augustin et al. 2012
	UGT73C12	*Barbarea vulgaris*	AFN26668	β-Amyrin, hederagenin	C3- β-D-glucosylation	Augustin et al. 2012
	UGT73C13	*Barbarea vulgaris*	AFN26669	β-Amyrin, hederagenin	C3- β-D-glucosylation	Augustin et al. 2012
	UGT73K1	*Medicago trunctula*	AAW56091	Hederagenin, soyasapogenol B/E	β-D-glucosylation	Achnine et al. 2005
	UGT73F3	*Medicago trunctula*	ACT34898	Hederagenin	C28-β-D-glucosylation	Naoumkina et al. 2010
	UGT73F2	*Glycin max*	BAM29362	Saponin A0-αg	C3'-β-D-glucosylation	Sayama et al. 2012
	UGT73F4	*Glycin max*	BAM29363	Saponin A0-αg	C3'-β-D-xylosylation	Sayama et al. 2012
	UGT73P2	*Glycin max*	BAI99584	Soyasapogenol B 3-*O*-glucuronide	C2'-β-D-galactosylation	Shibuya et al, 2010
74	UGT74M1	*Medicago truncatula*	ABK76266	Medicagenic acid	C28-β-D-glucosylation	Meesapyodsuk et al, 2007
91	UGT91H4	*Glycin max*	BAI99585	Soyasaponin III	C2" -β-D-rhamnosylation	Shibuya et al, 2010

1.4   Avenacins

1.4.1   Role in plant defence

Avenacins are monodesmosidic triterpene saponins found in oat roots. The ability to synthesise these compounds is restricted to the genus *Avena* (Fig. 1.7.A). Avenacins are antimicrobial and have been shown to protect oats against attack by soil-borne fungal pathogens, including the causal agent of take-all disease of cereals, *Gaeumannomyces graminis* var. *tritici* (Papadopoulou et al. 1999). Take-all disease symptoms are caused by blockage of the vasculature of the root by fungal hyphae, which leads to blackening of the root, pronounced leaf yellowing, bleached seed heads and reduced yield . The infection is easily spread from plant to plant by fungal hyphae growing under the soil surface (Cook 2003). Resistance to take-all has not as yet been found in cultivated wheat accessions. In the UK, 60% of wheat crops are estimated as being at risk to the disease with annual yield losses reaching £40-60 million. The introduction of a source of effective genetic resistance to take-all into cultivated wheat is expected to increase wheat yields by 10-50% in affected crops (http://www.takeall.com).

Figure 1.7: Avenacin structure and localization. A. Structures of the four avenacins: Avenacin A-1: R1 = OH, R2 = NHCH3; avenacin A-2: R1 = OH, R2 = H; avenacin B-1: R1 = H, R2 = NHCH3; avenacin B-2: R1 = H, R2 = H. The trisaccharide moiety of the avenacins consists of one L-arabinose linked to carbon 3 of β-amyrin and two D-glucose molecules linked to L-arabinose (β-1-4 and β-1-2 linkages). **B.** Localization of avenacin A-1 in oat root. Upper panel: Cross section of an oat root tip with nuclei stained with 4',6-diamidino-2-phenylindole (DAPI) showing the different cell types (e, epidermis; s, subepidermis; c, cortex). Bar = 50 μm. Lower panel: Cross section under ultra-violet (UV) illumination showing blue fluorescence attributable primarily to avenacin A-1 (Wegel et al. 2009).

The antimicrobial properties of avenacins are likely to be due to the ability of these molecules to permeabilise fungal membranes (Fig. 1.5). Like other saponins, avenacins are able to integrate into the outer monolayer of cellular membranes prior to complexing with sterols to cause membrane disruption. Incompletely glycosylated forms of avenacins are unable to form complexes with sterols, suggesting an essential role of the carbohydrate moiety of avenacin in avanacin-sterol interactions (Armah et al. 1999).

1.4.2 Avenacin structure

The major avenacin present in oat roots is avenacin A-1 but other variants exist, namely avenacins A-2, B-1 and B-2 (Fig. 1.7.A) (Crombie et al. 1984). Avenacins are oleanane-derived saponins, the apolar part of the molecule being composed of an oxidised form of β-amyrin (oleanane). One epoxide group, three hydroxyl groups and an aldehyde decorate β-amyrin to form 12,13-epoxy-16,21,23-hydroxy-30-oxo-β-amyrin. The hydroxyl group at the C-23 position is specific to B-type avenacins. The oleanane triterpene is acylated at C-21 with *N*-methylanthranilate (A-1 or B-1) or benzoate (A-2 or B-2). The *N*-methylanthranilate part of the molecule is responsible for the bright-blue fluorescence of avenacins observed in epidermal cells of oat root tip under UV illumination (Fig. 1.7.B) (Haralampidis et al. 2001).

The polar part of avenacin A-1 consists of a trisaccharide composed of one L-arabinose linked in the alpha configuration to C-3 of the β-amyrin skeleton, and two D-glucose (β-1-4 and β-1-2 linked) molecules linked to the triterpene via the L-arabinose (Crombie et al. 1984). Glycosylation of avenacin A-1 is essential for its antimicrobial activity. Some avenacin-resistant fungi have evolved a strategy to detoxify avenacin by hydrolysis of the terminal sugars using glycosidases. The enzyme avenacinase, which is produced by an avenacin-resistant relative of the take-all fungus, *Gaeumannomyces graminis* var. *avenae*, is able to remove the β-1,2- and β-1,4-linked terminal glucoses to give the less inhibitory bis-deglucosylated avenacin (Bowyer et al. 1995).

1.4.3 Biosynthesis of avenacins

Accumulation of avenacin in epidermal cell layers of the root cause bright blue fluorescence of oat root tip under UV illumination (Fig. 1.7.B). This property has been exploited to screen sodium azide-generated mutants of diploid oat (*Avena strigosa*) for avenacin-deficient (saponin-deficient) mutants. A set of 10 independent *saponin-deficient* (*sad*) mutants with reduced root fluorescence was identified in an initial screen (Papadopoulou et al. 1999). Gene cloning has shown that most of the *Sad* genes originally defined by mutation are linked in the oat genome (Qi et al; 2004) (Fig. 1.8.A). The avenacin gene cluster characterised to date contains the *Sad1, Sad2, Sad7, Sad9* and *Sad10* genes (Fig. 1.8.A) (Mugford et al. 2013). The uncloned *Sad3* locus is also physically linked to the gene cluster but further away on the oat genome (3.6 cM). The cloned *Sad* genes are all expressed in the epidermal cells of the root tips of *A. strigosa* seedlings (Fig. 1.8.B). Transcriptional regulators of avenacin biosynthesis have not yet been identified. *Sad10* is the only *UGT* gene present within the characterised part of the gene cluster.
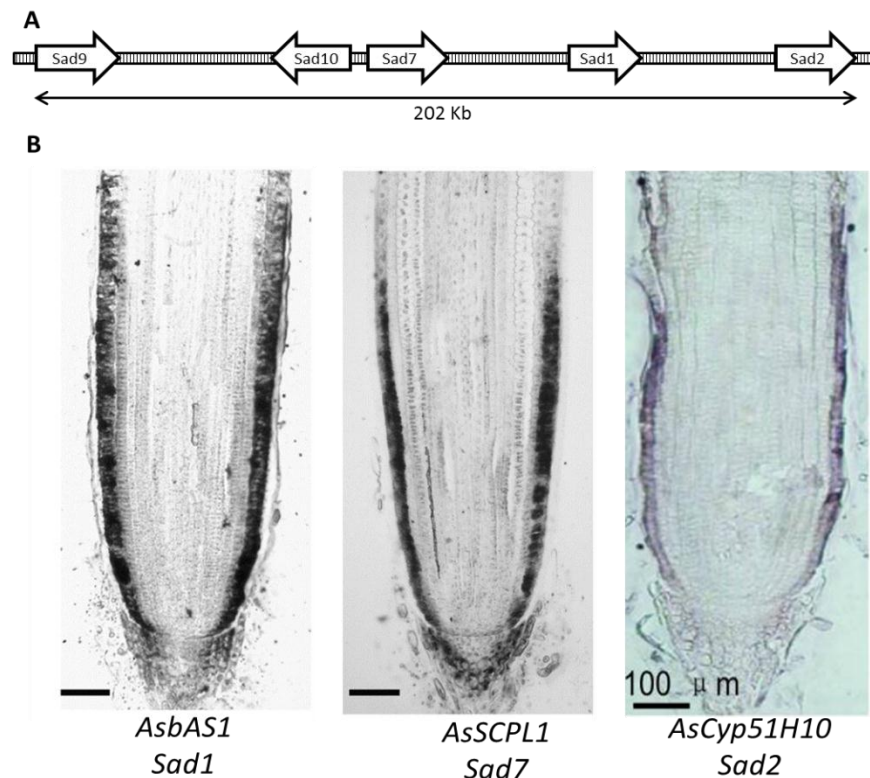
Figure 1.8: The avenacin biosynthetic gene cluster. A. Organisation of the avenacin gene cluster in the *A. strigosa* genome **B.** *In situ* mRNA hybridization showing the distribution of the *Sad1*, *Sad7* and *Sad2* transcripts in the roots of *A. strigosa* seedlings (Mugford et al, 2009).

Significant progress has been made in cloning and characterising the *Sad* genes involved in avenacin pathway (Fig. 1.9). Most of these genes reside within an 'operon-like' metabolic gene cluster in the oat genome (Qi et al. 2004). The first committed step in avenacin synthesis is catalysed by β-amyrin synthase (SAD1), an oxidosqualene cyclase that converts the ubiquitous precursor 2,3-oxidosqualene to β-amyrin (Haralampidis et al. 2001). The second step is catalysed by SAD2, a CYP450 that adds a hydroxyl group at C-16 of β-amyrin and also catalyses formation of a C-12/C-13 epoxide (Geisler et al. 2013). *Sad1* and *Sad2* have both been recruited from sterol metabolism by gene duplication and neofunctionalisation (Qi et al. 2006). SAD1 is a divergent member of the OSC family that is likely to have been recruited directly or indirectly from oat cycloartenol synthase (CAS1), which catalyses the cyclisation of 2,3-oxidosqualene to the sterol precursor cycloartenol (Haralampidis et al. 2001; Qi et al. 2004). The evolution of OSCs from cycloartenol synthases is common phenomenon in higher plants (Xue et al. 2012). SAD2 (CYP51H10) is a divergent member of the CYP51 sterol demethylase family (Geisler et al. 2013). Non-sterol-related genes have also undergone neofunctionalisation to form avenacin-biosynthesising enzymes. *Sad9*, *Sad10* and *Sad7* genes are required for the acylation of the avenacin A-1 and B-1 at position C-21 by the fluorophore *N*-methylanthranilate (Mugford et al. 2013). Anthranilate originated from the shikimate pathway is converted in *N*-methylanthranilate by the methyltransferases SAD9. *N*-methylanthranilate is further processed in *N*-methylanthranilate-*O*-glucose by the glycosyltransferase SAD10 (UGT74H5) (Owatworakit et al. 2012). The related enzymes UGT74H6 and UGT74H7 have been proposed to be involved in glucosylation of benzoate to generate the acyl donor required for synthesis of avenacins A-2 and B-2 (Owatworakit et al. 2012). The *Sad10* gene was first identified because it lies within the avenacin gene cluster. *Sad10* mutants have not yet been identified. *N*-methylanthranilate-*O*-glucose is used as an acyl donor by SAD7, a serine carboxypeptidase-like acyltransferase (AsSCPL1). SAD7 catalyses the transfer of *N*-methylanthranilate or benzoate onto C-21 of β-amyrin (Mugford et al. 2009).
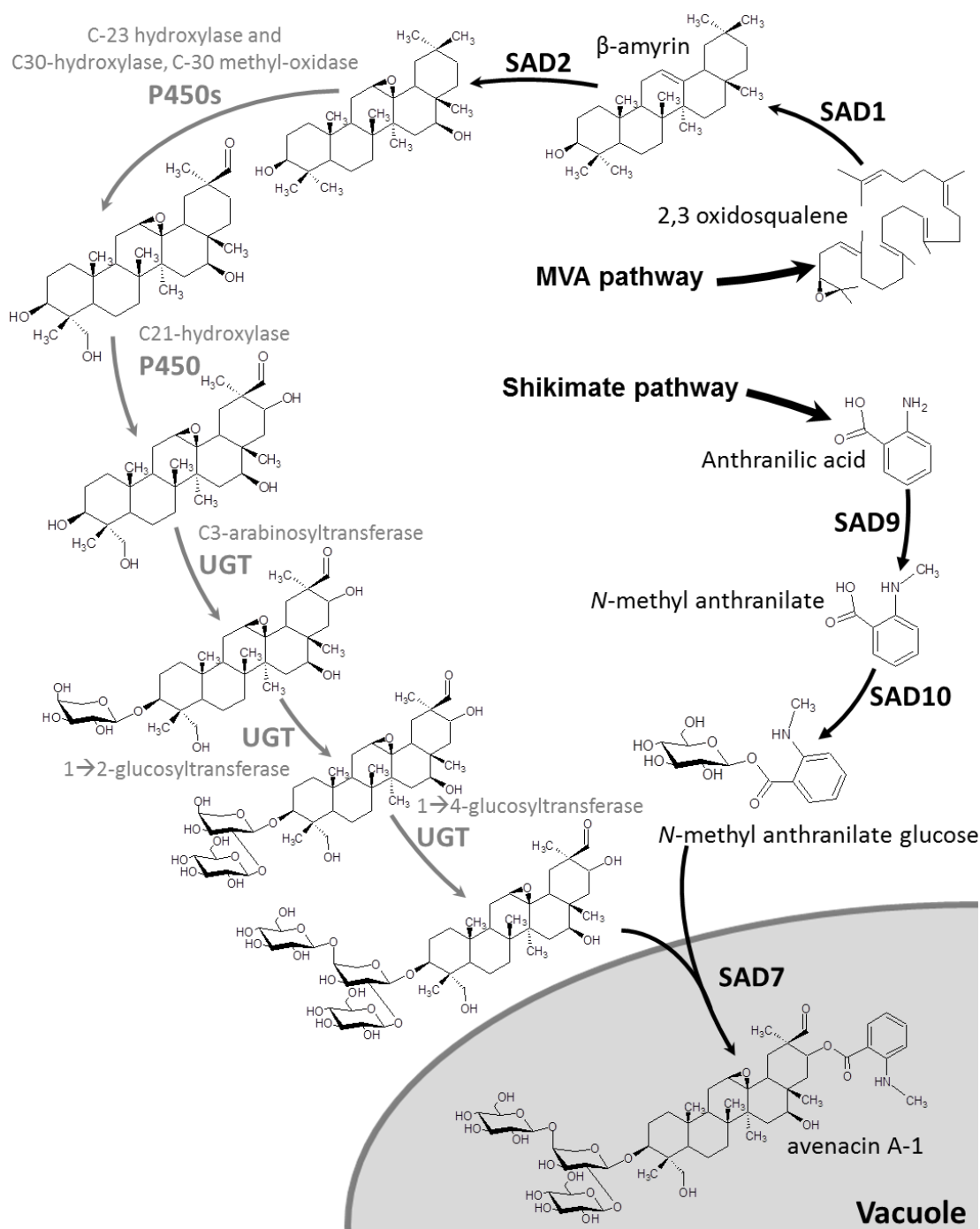
Figure 1.9: Current knowledge of the biosynthesis of avenacin A-1. Cyclisation of β-amyrin from 2,3 oxidosqualene is catalysed by β-amyrin synthase (SAD1). β-Amyrin is then modified by the cytochrome P450 CYP51H10 (SAD2) by both epoxidation and hydroxylation. The steps in grey are not known, a proposed order is shown based on *sad* mutants phenotypes. The fluorophore *N*-methyl anthranilate is generated from the shikimate pathway. Anthranilate is converted to *N*-methylanthranilate by the methyltransferase SAD9. This product is then glucosylated by the glycosyltransferase UGT74H5 (SAD10) to give *N*-methyl anthranilate glucose, which serves as the activated acyl donor for the serine carboxypeptidase-like acyltransferase (SAD7) in the vacuolar compartment. Steps unknown to date are shown as hypothetical reactions with grey arrows.

Although substantial progress has been made in cloning and characterising the genes and enzymes for avenacin biosynthesis the part of the pathway that is concerned with the synthesis of the trisaccharide sugar chain remains poorly understood. Mutants that are affected at two as yet uncloned loci – *Sad3* and *Sad4* – have been shown to accumulate the same incompletely glycosylated avenacin intermediate, namely avenacin lacking the *β*-1-4 linked glucose (Fig. 1.10.A) (Mylona et al. 2008). Unlike the other avenacin deficient mutants that have been isolated, these mutants have root morphology defects (Fig. 1.10.B-D). At the cellular level the *sad3* and *sad4* mutants appear to have membrane trafficking defects that lead to degeneration of the epidermal cell layer (Fig. 1.10.E-G). It is unclear whether the *Sad3* and *Sad4* gene products are directly involved in avenacin biosynthesis (for example, in glycosylation, transport or regulation) or whether they have a broader function (e.g. in vesicular trafficking). It is interesting to note that the *sad1* mutation when introduced into a *sad3* or a *sad4* mutant background suppresses the root morphology phenotype. *sad1* mutants are blocked in the first committed step in avenacin biosynthesis (the cyclisation of 2,3-oxidosqualene to β-amyrin) and so *sad1/sad1 sad3/sad3* and *sad1/sad1 sad4/sad4* lines do not accumulate the monodeglucosylated avenacin intermediate. Therefore the detrimental effect on root morphology seen in *sad3* and *sad4* mutants is likely to be due to accumulation of the monodeglucosylated avenacin intermediate inside or just outside the vacuole. The *Sad3* locus is genetically linked to the avenacin cluster (Qi et al. 2004). In contrast, the *Sad4* locus is not linked to the avenacin biosynthetic gene cluster and *sad4* mutants are also affected in glycosylation of the steroidal saponin avenacoside found in oat leaves, pointing towards a broader function for the *Sad4* gene product. Supporting this, the accompanying root phenotype of *sad4* mutants is less pronounced than *sad3* mutants.
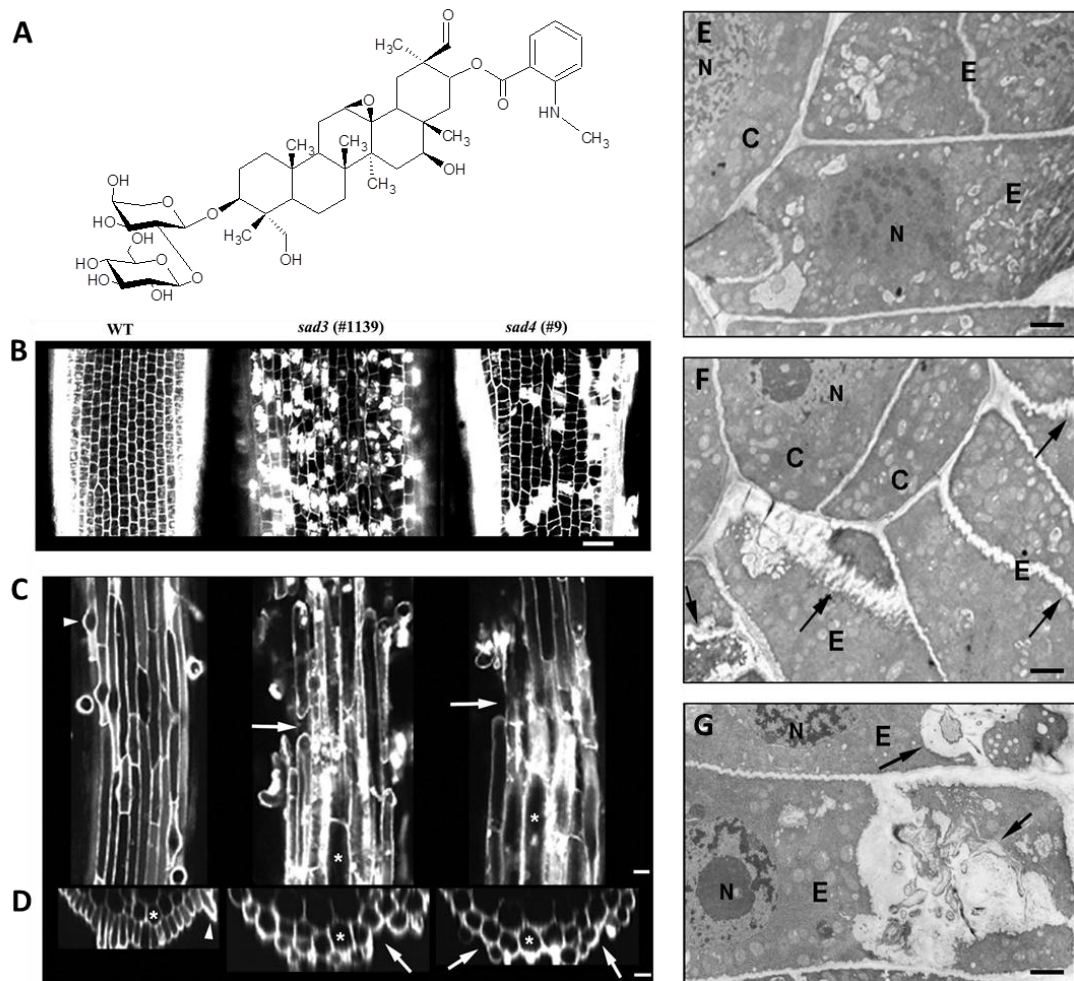
Figure 1.10: Phenotypes of *sad3* and *sad4* mutants (Mylona et al. 2008). A. *sad3* and *sad4* mutants accumulate avenacin A-1 lacking the β-1,4-linked glucose. **B-D**. Degeneration of the epidermal cell layer in *sad3* and *sad4* mutants. Bars = 25 μm. (C) Longitudinal optical sections of roots of 2 day old wild-type and mutant propidium iodide–stained seedlings, showing the epidermal cell layer in the meristematic zone. **E-G.** *sad3* and *sad4* mutants have membrane trafficking defects. Transmission electron micrographs of cross sections of the epidermal cells of roots of 2 days old wild-type (E) and sad3 mutant ([F] and [G]) seedlings in the meristematic zone. Bars = 2 μm (E), 6 μm (F) and 1.8 μm (G).

Apart from *sad3* and *sad4* mutants no other oat mutants affected in glycosylation of avenacin have been identified (Papadopoulou et al. 1999). Avenacin glycosylation is likely to be mediated by UGTs. UGTs are the only enzymes reported in saponin glycosylation to date (Augustin et al. 2011). Considering the regiospecificity of UGTs, the sequential synthesis of the avenacin trisaccharide is likely to be performed by three different UGTs (Fig. 1.11). However, we cannot exclude the hypothesis of a single enzyme catalysing the two glucosylation reactions regarding the few examples of UGTs catalysing consecutive glycosylation of the same acceptor. Such activities have been reported for RhGT1 of

Rosa hybrida synthetizing cyanidin-3,5-O-diglucoside from cyanidin (Ogata et al. 2005) or UGT73C12 and UGT73C13 of *B. vulgaris* catalysing formation of bidesmosidic diglucosides of oleanane triterpenes *in vitro* (Augustin et al. 2012).
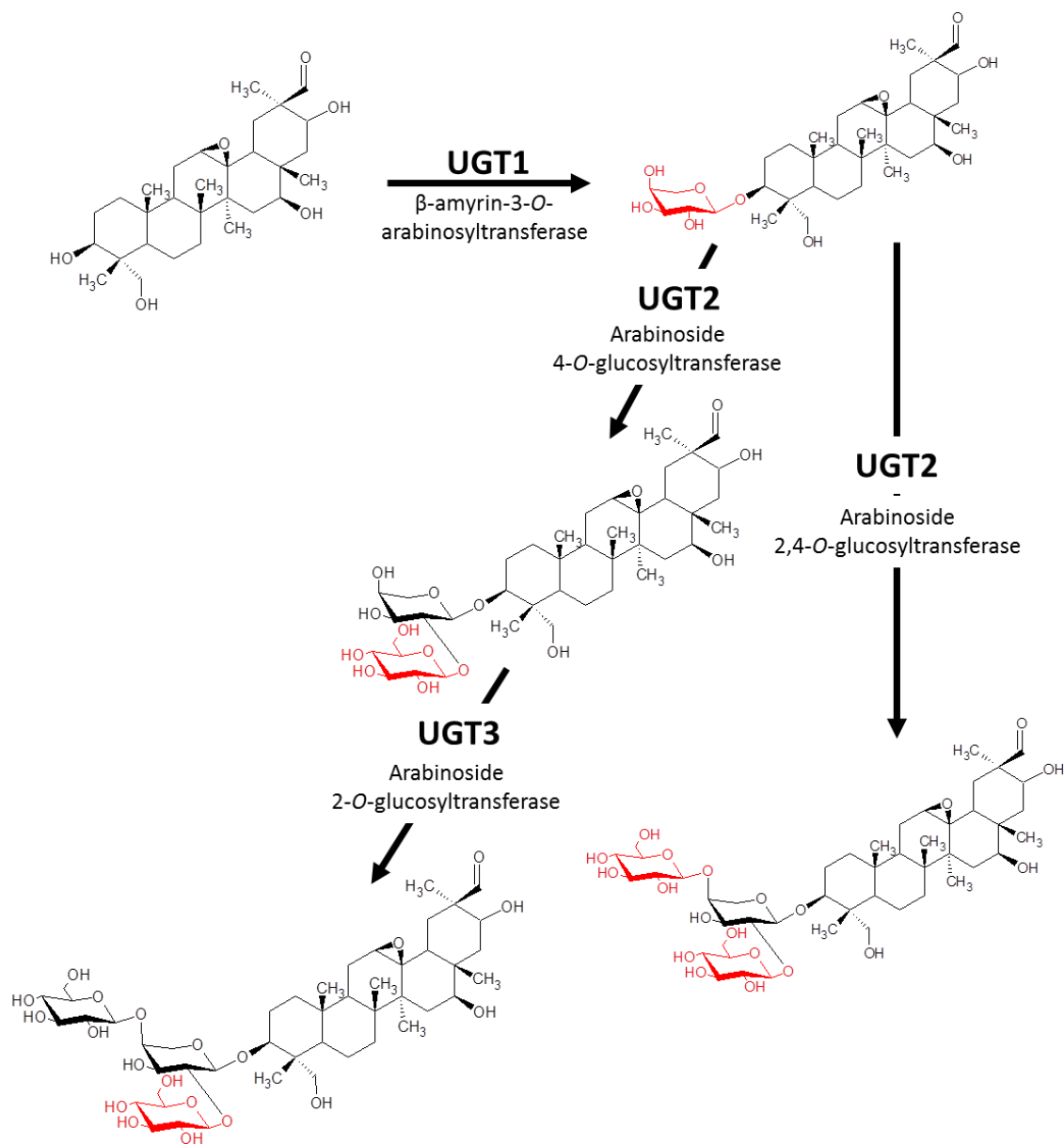
Figure 1.11: Hypothetical model for the formation of the avenacin trisaccharide. Glycosyltransferases belonging to family 1 (UGTs) are known to be involved in saponin glycosylation in other plant species and are therefore obvious candidates for synthesis of the trisaccharide moiety of the avenacins. Two possible pathways are proposed here, involving either two or three UGT enzymes. A specific UGT is likely to add arabinose onto the hydroxyl group of carbon 3 of β-amyrin. Following this first step, two glucosylation events onto carbons 2 and 4 of the arabinose molecule are required to complete the formation of the trisaccharide. These two glucose molecules may be added either by a single enzyme or by two regiospecific glucosyltransferases.

1.5   PhD project overview

Glycosylation of plant natural products is essential for cellular homeostasis, defence against biotic and abiotic stresses, and xenobiotic detoxification. Despite the growing body of literature on glycosyltransferases of plant natural products, the features that determine sugar acceptor specificity and regioselective addition of sugars are still only poorly understood. In the past, considerable effort has been invested in characterising glycosyltransferases from the model plant *A. thaliana* and other dicot species (e.g. *M. truncatula*, *Lamiales*). Comparatively, only a few UGTs have been characterised from monocots, mainly from rice and maize. A systematic analysis of monocot family 1 glycosyltransferases will provide important insights into the function and evolution of natural product glycosyltransferases in higher plants. The roots of plants are rich sources of natural products. In oats these include the antimicrobial triterpene glycoside avenacins. Here a systematic analysis of UGTs that are expressed in the roots of diploid oat (*A. strigosa*) is undertaken to develop a better understanding of the enzymes that glycosylate small molecules, including avenacins, in monocots. Investigation of these oat UGTs also has the potential to generate useful biocatalysts for the synthesis of glycoconjugates of biotechnological interest. The aims of this project were therefore to:

- Take a systematic approach towards identification and analysis of UGTs expressed in oat roots.

- Identify avenacin glycosyltransferases candidates and functionally characterised to determine their potential role in the avenacin pathway.

- Develop *E. coli* and plant-based systems for expression and functional analysis of oat UGTs and establish an *in planta* synthetic biology platform for triterpene engineering.

- Develop novel assays for UGT activity and investigate the activity oat UGTs toward terpenoids and other potential acceptors of biotechnological interest.

Chapter 2 - Materials & methods

2.1   <u>Materials</u>

2.1.1   <u>Biological materials</u>

Oat plants used in this study are *Avena strigosa* accession S75 (from the Institute of Grasslands and Environmental Research, Aberystwyth, Wales, United Kingdom). Conditions of growth are described in section 2.2.1. Saponin-deficient mutant lines of *A. strigosa* originate from sodium azide mutagenesis (Papadopoulou et al. 1999). In this thesis, mutant lines from M5 progeny #1139 (*sad3* mutant line) and #9 (*sad4* mutant line) (Mylona et al. 2008) were used. *Nicotiana benthamiana* plants were grown in green-houses maintained at 23°C to 25°C with 16 hours of supplementary light per day.

2.1.2   <u>Bacterial strains</u>

species	strains	Supplier/origin	Use
Agrobacterium tumefaciens	LBA4404		Transient expression in *N. benthamiana*
Escherichia coli	DH5α	Invitrogen	Gateway cloning
	BL21	Invitrogen	Protein expression
	BL21 Rosetta	Invitrogen	Protein expression
	BL21 Lemo		Protein expression

2.1.3   Primers

Race PCR primers:

Name	Sequence	Annealing temperature
GeneRacer™ 5′ Primer (Invitrogen)	CGACTGGAGCACGAGGACACTGA	68.6
GeneRacer™ 5′ Nested Primer (Invitrogen)	GGACACTGACATGGACTGAAGGAGTA	56.9
F-M13	GTTTTCCCAGTCACGAC	51.5
R-M13	CAGGAAACAGCTATGAC	83.8
R1-27f7	GGGAGAGGGTCTCCAGCGTGCCGTTCCA	84.1
R2-27f7	GCCCTGTCCTTGACCCGCGCCTCGAA	83.3
R1-3i21	GCCATCAGCGCCCGTGCCACAGCTT	83.5
R2-3i21	CGTCTCCAGCACCGAGTTCCACCCGCAGT	85.2
Rr-3i21	GCAGGCGGCCTGCGCCGCGTAGAA	72.2
Rn-3i21	CGTGGAGTAGTCCACCTCCTGCTTCT	82.4
R1-14h20	CCGGCATGCCTGGCACCTCCACGTA	69.4
R2-14h20	CACAGTGAGTAGAAGCAGGAGGCTACAT	76
Rn-25n16	CATTTCAGGGTAGCTGTGGGGCAGGTT	84.1
Rr-25n16	GCCTGGGAAGCGCACGAGCGCTTGGT	81.3
Rn-5d1	GGTGGGACGCCGAGAGCTCGTCCA	88.5
Rr-5d1	TCCCCCGCCGCGAACGCCGCCAT	81.8
R1-8i4	CGCCCCTCCGTCTCCGCCGAGAATT	78.4
Rn-8i24	GCACGTTCCAGAGGAAGGCGTAGCCTGT	68.6

Transcripts expression analysis primers for RT-PCR:

Name	Sequence	Annealing temperature
Fep-1a15	GCGCCCCACCTGCTACTGATAT	70.3
Rep-1a15	TTCCCGACGACGCAAGACACAG	73.8
Fp-3i21	CAGCCGAAAGCACCATCCC	70.4
Rp-3i21	TGGAGTAGTCCACCTCCTGCTTC	68
F-8i4	GAGCATGCTGTCCACGTGGT	69.2
R-8i4	CCGTTTTAGTCCATGCAAGC	64.7
Fp-11i11bis	GATGATGATGATGCAGTGGTGGAG	70
Rp-11i11bis	TGCTGCTAGATATTGGCGGC	67.5
F-14h20	AACTTCCTCCGGCTGTTCCT	66.5
R-14h20	CGCACGAAGCTATGCACTAG	64
F-14h21	ATATCAGTTGATTCACCGTTGTTT	62.6
R-14h21	AGAGCTGTCCGGTAGCACG	66
F-15a11	CTGACGGATTCAGCGAAAAA	65.4
R-15a11	AGACCGTAGTGCATGAAGAAAT	61.9
Fp-16f23	GGCTTCTGTGGCTTCTCGTCC	69.7
Rp-16f23	TGTGTTGCTGTCTCGGTGGCA	72.9
Fp-16h6	GGTGGTTGACGGAGAGGAGGA	70.1
Rp-16h6	CCGCCTTCCACGATAGCTGTTT	70.3
F-20e13	GTGTGGGTACCGCTGCATAT	65.1
R-20e13	CGCCTCTAGCTCATCCCCAT	67.9
Fep-21p16	ATGGGGTCGATGGAGCAGAA	69.3
Rep-21p16	TAGGCGAACGACATGCAGGC	70.9

Fep-24a3	GAGCGCGTCGAATCAGGGC	72.9
Rep-24a3	CCGTGCGCCAGGATCTCCAG	74.8
Fp-24i2	ACGCCGCCGACCAGAAGAG	72.4
Rp-24i2	CCAAGCACCCACACGAAATTC	68.9
Fp-25n16	GCCCCACAGCTACCCTGAAATG	70.8
Rp-25n16	CTCTTTTCGCCGTAGCCGC	69.3
Fp-27a12	ATATTCATGGCGGAGGACGGCG	74.5
R-Nes27a12	GCTTTTGCTCGACGGGCCGC	76.9
Fep-27f7	GTTTCGAGGCGCGGGTCAAG	73.3
Rep-27f7	CATCAACTCGGCGACCGCCT	74
F-28b19	TCGACGACGGCACCAACGCT	75.7
R-28b19	ACGGCCGAGGACCCCATCAC	74.9
F-30a8	GACATCGTCGCGCAGCTCCT	72.6
R-30a8	GCGTCCCTTCCGTGCCATGG	76.2
Fp-00733	CAGTCCGTGATCTTCGTCGC	68.6
Rp-00733	CCCACTCTTAACCAACCCTCCAC	68.7
Fep-01989	GGCCACATCAACCCGATCCTG	73.1
Rep-01989	CCAGCCCAGGCGTGCTCG	75.3
Fp-02436	CTCTACAGGCAAGGGCGGCA	72
Rp-02436	CTCCTTGCCATCCAGCCAAG	69.4
Fp-03999	AGCGGGAGAGGGAAGCGGA	72.9
Rp-03999	GGCGATAAGGGCACTACGTGTC	68.9
Fp-05827	CGATGGACAGTAGTGTGGCGTT	68.5

Rp-05827	GCGGTGGGTGGCCGAGAG	74.2
Fp-10319	GCGTGAAGATCCCTGCCAT	67.9
Rp-10319	GTCTTCAATCTTCCCGGTGTCC	68.1
Fp-11637	AGGACACGGGCTGTCTTCAGT	67.7
Rp-11637	TGACGAAGCACGCCACGG	72.6
Fep-16327	CGCAACAGGGCCACCGTC	72.9
Rep-16327	TGCATCACCAGCCTCGGG	71.5
Fep-16525	GCCCTACCGTCGTCCTCATT	67.6
Rep-16525	AAGAGATCCACGACGACCGC	69.1
Fep-23586A	GGAAGACGAGATGTGCAAGATCC	68.3
Rep-23586A	TCCAAGTATTTCTTGGCGTCCTG	68.3
Fep-23586B	GAGGAGCTGTGCAAAGAAATCAC	66.4
Rep-23586B	GTCCGCTCTAGGAAACCTTCTGG	68.2
Fp-24138	GTGGAAGAGGTAGGAGGACCG	66.3
Rp-24138	CAGGATCATTATTATTTGTCGACG	64.4
Fp-23781	TGCTGACCCGAAACAGATCATG	69.6
Rp-23781	CCACCACGATCTCCTTCTTGTC	67.3
Fp-24525	AGGAGGAGGAGGAGGAGCGGC	73.3
Rp-24525	CAAGCGGAGCCAAATCCATACTGC	73
Fp-SGT1	CGGCGGCACTACCCACGG	74.5
Rp-SGT1	AGGATCATTATTCTGCCGGCG	69.3

Gateway cloning primers:

Name	Sequence	Annealing temperature
attB1 adapter primer	GGGGACAAGTTTGTACAAAAAAGCAGGCT	72.9
attB2 adapter primer	GGGGACCACTTTGTACAAGAAAGCTGGGT	74.7
attL1	CAAATAATGATTTTATTTTGACTGATAGT	60
attL2	CTATCAGTCAAAATAAAATCATTATTTG	59.4
Fgw-Sad10	AAAAAGCAGGCTTAATGGGGGCTGAGTGGGAGCA	82.2
Rgw-Sad10	AGAAAGCTGGGTATCATGCATCTAACCCCACCAGCA	80.9
Fgw-AsGT3i21	AAAAAGCAGGCTTAATGGCCTCTACCACCACCGCTA	80
RgwAsGT3i21	AGAAAGCTGGGTATCAGCCTTGAACTGACTTGGGCC	80.7
Fgw-11i11	AAAAAGCAGGCTTAATGGCCTCTAACGATAATGTCCCCACGG	82.1
Rgw-11i11	AGAAAGCTGGGTATCAGGTCCGCGGCCTCTTTGCTG	84.4
Fgw-14h20	AAAAAGCAGGCTTAATGGCGCACACAGAGACGAC	80.1
Rgw-14h20	AGAAAGCTGGGTATCAACAATTCCGATCTTGGGTAG	76.7
Fgw-14h21	AAAAAGCAGGCTTAATGGCGCCCACGGAGACGGC	85.3
Rgw-14h21	AGAAAGCTGGGTATCACTGCTCCTTGCTGCCACTCCGC	84.9
Fgw-16f23	AAAAAGCAGGCTTAATGACCTTCGCCCGCGGC	83.1
Rgw-16f23	AGAAAGCTGGGTATCAGCCACATGCATTTGTG	77.7
Fgw-16h6	AAAAAGCAGGCTTAATGGCATCGAGGCAGTACCA	79.2
Rgw-16h6	AGAAAGCTGGGTATCACTTCCTCTGGGACGGGATCA	81.3
Fgw-21p16	AAAAAGCAGGCTTAATGGGGTCGATGGAGC	77.4
R2gw-21p16	AGAAAGCTGGGTATCAATTTCGGCCAGTAGAT	74.1
Fgw-24i2	AAAAAGCAGGCTTAATGGCTGTCAAAGATGAGCA	77.1

Rgw-24i2	AGAAAGCTGGGTATCAAAGTCCGTCATTTGTCGGGA	80.2
Fgw-27a12	AAAAAGCAGGCTTAATGAAGGACGCGACGGCGAC	82.3
Rgw-27a12	AGAAAGCTGGGTATCACCTGCTGGTGATGTAAGCAA	78.8
Fgw-25n16	AAAAAGCAGGCTTAATGGCAGGCATGGCTCCGCT	83
Rgw-25n16	AGAAAGCTGGGTATCAAGTGCTTGTTGAATTCTCCA	76.9
Fgw-27f7	AAAAAGCAGGCTTAATGGGGACGTTGTCGGAGCT	80.3
Rgw-27f7	AGAAAGCTGGGTATCAAGACTGTACTGACAGTGCAG	75.4
Fgw-02436	AAAAAGCAGGCTTAATGGAGACCTCCGCAA	76.6
Rgw-02436	AGAAAGCTGGGTATCACGCACAAGAATCGATC	76.4
Fgw-03999	AAAAAGCAGGCTTAATGGGGTACAACGGCG	77.5
Rgw-03999	AGAAAGCTGGGTATCAAAAATTGAATGGGAGA	73.3
Fgw-05827	AAAAAGCAGGCTTAATGGGGATTGAGTCGATGGACA	79.9
Rgw-05827	AGAAAGCTGGGTATCAAATCCTTGTGATGTGAGCAA	77.6
Fgw-16525	AAAAAGCAGGCTTAATGGCGCCTCGCCCTACCGTCGT	85.2
Rgw-16525	AGAAAGCTGGGTATCACCTACCAACCGGCAAATCCA	81
Fgw-23586	AAAAAGCAGGCTTAATGAAGCAGACCGTCGTCCTGT	79.6
Rgw-23586	AGAAAGCTGGGTATCACTCGCGTACCTGCTCTCCGA	81.9
Fgw78D3	AAAAAGCAGGCTTAATGGCCAAACCCTCGCAG	79.7
Rgw78D3	AGAAAGCTGGGTATCATTATCCAAAGTTCACAACTT	72.7
Fgw73C10-12	AAAAAGCAGGCTTAATGGTTTCCGAAATCACC	75.2
Rgw73C10	AGAAAGCTGGGTATCATCAATTATTAGGTTGTGC	71.9
Rgw73C12	AGAAAGCTGGGTATCATCAATTATTGGATTGTGC	73.9

RT-PCR primers for pEAQ transient expression:

Name	Sequence	Annealing temperature
F-5′ UTR pEAQ	ACTTGTTTGATCGAATTTGG	53.7
F-pNos	GAACTGACAGAACCGCAACG	60.3
R-Kan	TCAGCAATATCACGGGTAGC	57.2
F-SAD1	ATGTGGAGGCTAACAATAGG	58.1
R-SAD1	TATCTCATGACGATGTTCCG	61.1
F-SAD2	ATGGACATGACAATTTGCGT	57.3
R-SAD2	CATTGGCACGGTGAACTCAT	65.9
Fep-UGT73C10	GTTTCCGAAATCACCCATAA	55.9
Rep-UGT73C10	AAGACAAAAGCAACACATGC	54.9

2.1.4 Plasmids

Plasmids name	Supplier/origin	Use
pDONR207	Invitrogen	Entry clones
pCR4-TOPO	Invitrogen	Cloning of PCR product
pH9-GW	O'Maille laboratory, gateway-compatible version of pET18	Expression in *E.coli* with 9xHis N-term tag
pEAQ-HT-DEST1	Lomonossoff laboratory	Transient expression in *N.benthamiana*
pEAQ-HT-DEST2	Lomonossoff laboratory	Transient expression in *N.benthamiana* with 6xHis N-term tag

Plasmid maps are presented in chapter 4, figures 4.2 and 4.3

2.1.5 Chemicals

Chemicals	formula	Supplier/origin	Purity
2,4,5-Trichlorophenol	$C_6H_3Cl_3O$	Sigma-Aldrich	95%
Uridine diphospho-β-D-glucose	$C_{15}H_{23}N_2O_{17}P_2$	Sigma-Aldrich	98%
Uridine diphospho-β-D-galactose	$C_{15}H_{23}N_2O_{17}P_2$	Sigma-Aldrich	98%
Uridine diphospho-α-L-arabinose	$C_{14}H_{21}N_2O_{16}P_2$	Carbosynth Ltd	98%
Uridine diphospho-β-D-glucose [6-^{3}H]	$C_{15}H_{23}N_2O_{17}P_2$	ARC Inc	-
Quercetin	$C_{15}H_{10}O_6$	Sigma-Aldrich	$\geq 95\%$
Kaempferol	$C_{15}H_{10}O_6$	Sigma-Aldrich	$\geq 97\%$
Tricin	$C_{17}H_{14}O_7$	SelectLab Chemicals	95%
β-amyrin	$C_{30}H_{50}O$	Apin Chemicals Ltd	-
β-amyrin arabinoside	$C_{35}H_{58}O_5$	Field Lab, chemical synthesis	-
Avenacin A-1	$C_{55}H_{80}NO_{21}$	Purified from *A. strigosa*	-
Hederagenin	$C_{30}H_{48}O_4$	Apin Chemicals Ltd	-
Lupeol	$C_{30}H_{50}O$	Apin Chemicals Ltd	-
Capsidiol	$C_{15}H_{24}O_2$	O'Maille Lab, purified from *Capsicum annuum*	99%
α-bisabolol	$C_{16}H_{28}O_2$	Sigma-Aldrich	$\geq 95\%$

2.2   Methods

2.2.1   Growing conditions of *A. strigosa* and tissue collection

Appropriate number and variety of seeds were dehusked manually, sterilised in 5% sodium hypochlorite solution and shaken for 5 min (Spiramix 5, Denley). Seeds were rinsed 5 times in sterile water. Dried seeds were placed onto Petri dishes containing distilled water agar (15 seeds per plate). Plates were stored in dark at 4°C for 12 hours for synchronisation of germination. Seeds were transferred to a growth cabinet (SANYO, Versatile Environmental Test Chamber) (22°C, 16 hours of light per day). After 3 days of growth, young tissues were harvested. *A. strigosa* root tips (last 0.5cm), elongation zone, whole root and/or whole leaves were harvested with a sterile razor blade and frozen directly in liquid nitrogen. *A. strigosa* mature tissues were harvested from flowering plants grown in green house conditions. Stalks, flowers, and/or mature leaves were harvested manually and frozen directly in liquid nitrogen.

2.2.2   Genomic DNA isolation from *A. strigosa*

A total of 250 mg of 3-day-old *A. strigosa* leaf tissues were ground and homogenised in liquid nitrogen using a pestle and mortar. Ground tissue was resuspended in 500 µL of DNA extraction buffer (50 mM Tris-HCl pH8, 100 mM NaCl, 50 mM ethylenediaminetetraacetic acid (EDTA), 10 mM dithiothreitol (DTT, Sigma-Aldrich), 1% Sodium dodecyl sulfate (SDS, Sigma-Aldrich)). The sample was vortexed and incubated 30 min at 50°C. DNA was extracted from the cell lysate twice using 500 µL of phenol/chloroform/isoamylalcohol (25:24:1, Sigma-Aldrich). DNA was precipitated from the aqueous phase by the addition of 50 µL of 3M sodium acetate and 1 mL of absolute ethanol. DNA was collected by centrifugation for 10 min at 14,000 g and the obtained DNA pellet was washed with 1 mL 70% ethanol before being resuspended in Milli-Q H_2O. Final concentration of the DNA solution was measured by Nanodrop (NanoDrop 8000 spectrophotometer, Thermo Scientific).

2.2.3   RNA isolation

Frozen tissues were ground in liquid nitrogen with an autoclaved mortar and pestle. 100mg of the tissue powder was homogenised in TRI reagent (Sigma-Aldrich) for 10 minutes at room temperature. Chloroform (0.2 mL) was added to the sample. The sample was shaken and centrifuged (12,000 g, 15 min at 4°C) to allow separation of the three layers. The aqueous upper phase was isolated, the addition of 0.5 mL of isopropanol and 50 µg/mL of glycogen allowed precipitation of RNA after centrifugation (12,000 g, 10 min at 4°C).

The dry pellet was resuspended in 43 µL of diethylpyrocarbonate (DEPC) treated water and DNA contaminants were digested by treatment with 2 µL DNase (Deoxyribonuclease I, Sigma-Aldrich) buffered with 5 µL of 10x reaction buffer. Treated RNA was extracted with 0.2 mL phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma-Aldrich) then concentrated by addition of sodium acetate (20 µL at 3 M, pH 5.2) with 50 µg/mL of glycogen and 520 µL of 95% ethanol. After centrifugation (14,000 g, 20 min at 4°C) and removal of supernatant, the RNA pellet was washed again with 75% ethanol. The RNA pellet was air-dried and resuspended in a suitable volume of DEPC-treated water and stored at -80°C.

The RNA concentration was measured using a NanoDrop 8000 spectrophotometer (Thermo Scientific) and the A260/A280 ratio was verified. RNA integrity was determined by separation of total RNA on denaturing agarose gel electrophoresis (see section 2.2.5) to check presence of 25S/18S RNA bands and distribution of mRNA.

2.2.4   cDNA synthesis

In an RNase-free tube 5 µg of RNA were mixed with 500 ng of poly-thymidine (oligo(dT) 15 primer, Promega) and 1 µL of dNTP mix at 10 mM in a final volume of 13 µL of DEPC-treated water. To avoid formation of RNA secondary structures, the sample was incubated for 5 min at 65°C and placed on ice for 1 min. Superscript III® kit (Invitrogen) components were added as follows: 4 µl of RT First-Strand buffer, 1 µl DTT 0.1 M, 1 µl of RNaseOUT 40 U/µl and 1 µl of SuperScript III reverse transcriptase 200 U/µl. Samples were incubated at 50°C for 45 min and the reaction was inactivated at 70°C for 15 min. RNA contaminants were

removed by RNase treatment at 37°C for 15 min. Concentration of single-strand cDNA was analysed by a Nanodrop 8000 spectrophotometer (Thermo Scientific).

2.2.5   DNA electrophoresis

The required quantity (1 g per 100 ml, for 1% agarose gel) of agarose powder was weighed out. The agarose powder was dissolved in TRIS-acetate EDTA (TAE) buffer (40mM Tris, 20mM acetic acid, and 1mM EDTA) using a microwave oven at full power until boiling. Ethidium bromide was added to cooled liquid agarose solution at a concentration of 500 µg/mL. The solution was poured onto the gel plate with combs used to form wells in the desired positions. After cooling, the gel was transferred to the electrophoresis tank and covered with TAE buffer.

Wells were loaded with 5 µL samples (1x Green buffer as a dye (GoTaq buffer, Promega), 1-2 µL of PCR product). DNA ladder (2 µL) was added onto one well (1x Green buffer, 100 µg/mL 1 kb DNA Ladder (New England Biolabs)). A current of 100 volts was applied to the gel for 20-25 min. Migration of DNA fragments was analysed under UV light using a gel imager (Gel Doc XR+ system, Bio-Rad). PCR products < 800bp were analysed on a 1.5% agarose gel.

2.2.6   Sequencing reaction of DNA fragment

Sequencing reactions were set up as follows:

Volumes	Mix component	Final concentration
x µl	DNA material	50-100 ng
2 µl	5x Big Dye Terminator V3.1 sequencing buffer (Invitrogen)	1x
0.5 µl	DMSO 100%	5%
2 µl	Big Dye V3.1	-
1 µl	Sequencing primer (10 mM)	1 µM
Up to 10 µl	sterile Milli-Q water	-

PCR program was performed as follows:

PCR steps	Duration	Temperature	Cycles
Denaturation	1.5 min	95°C	x1
Denaturation	20 sec	95°C	
Annealing	20 sec	50°C	X30
Elongation	4 min*	60°C	
Annealing	30 sec	50°C	x1
Elongation	4 min	60°C	x1

Analysis of the reaction products was performed by TGAC (Norwich Research Park). The sequences were analysed using Vector NTI Advance 11 software (Invitrogen).

2.2.7 Transcript expression analysis

cDNA library from *A. strigosa* tissues (see section 2.2.4) were diluted down to 100 ng/µL based on Nanodrop quantification (NanoDrop 8000 spectrophotometer, Thermo Scientific). Fragment amplification of constitutively expressed glucose-6-phosphate dehydrogenase (G6PDH) was used in the normalisation strategy in order to obtain similar level of amplification of G6PDH from each tissue. PCR mixes were prepared as follows:

Volumes	Mix component	Final concentration
1 µl	cDNA from *A. strigosa* tissue (100 ng/µL)	50 ng
4 µl	GoTaq 5x Green buffer (Promega)	1x
0.5 µl	dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP)	250 µM each
2 µl	MgCl$_2$ (25 mM)	2.5 mM
1 µl	Forward primer (10 µM)	0.5 µM
1 µl	Reverse primer (10 µM)	0.5 µM
0.2 µl	GoTaq DNA polymerase (5 U/µl, Promega)	1 U
9.5 µl	sterile Milli-Q water	-

PCR program was performed as follows:

PCR steps	Duration	Temperature	Cycles
Denaturation	2 min	95°C	x1
Denaturation	30 sec	95°C	
Annealing	30 sec	Var*	X30
Elongation	2 min*	72°C	
Elongation	10 min	72°C	x1

*annealing temperature is variable depending on primers used (see section 2.1.3)

PCR products were analysed by electrophoresis, 1.5% agarose gel (see section 2.2.5).

To control the transient expression of transcripts in *N. benthamiana* using the pEAQ system, cDNA were generated using 5'UTR specific primers. The DNA contamination experiment consists in amplifiying untranslated region of the T-DNA using F-pNos and R-Kan primers. The primers used in this analysis are listed in section 2.1.3.

2.2.8   Rapid amplification of cDNA 5'end (RACE PCR)

Amplification of *A. strigosa* UGTs 5'end was performed by RACE-PCR using GeneRacer Kit (Invitrogen).

Preparation of *A. strigosa* root tip cDNA with 5'end adapters

Dephosphorylation of non-mRNA contaminants was performed at 50°C for 1h. The reaction mix was composed of 10 units calf intestinal phosphatase (CIP), 1x CIP buffer, 40 units of RNaseOut and 5 µg of total RNA extracted from *A. strigosa* root tip (see section 2.2.3) diluted in DEPC-treated water. RNA was precipitated using the phenol:chloroform method (see section 2.2.3). Agarose gel electrophoresis was performed to assess RNA integrity (see section 2.2.5).

Decapping reaction of dephosphorylated RNA was performed at 37°C for 1h. The reaction mix was composed of 0.5 units of tobacco acid pyrophosphatase (TAP), 1x TAP buffer, 40 units of RNaseOut and 4.3 µg of dephosphorylated RNA diluted in DEPC H_2O. RNA was precipitated into phenol:chloroform (see section 2.2.3). Agarose gel electrophoresis was done to control RNA integrity (see section 2.2.5).

RNA secondary structure was denatured by incubation at 65°C for 5 min using 3.7 µg decapped RNA and 0.25 µg GeneRacer RNA Oligo (confer 2.0). The following components were added to 10 µL of reaction mix: 5 units of T4 RNA ligase, 1x ligase buffer, 1 mM of adenosine triphosphate (ATP), 40 units of RNaseOut diluted in DEPC H_2O. Ligation reaction was incubated 37°C for 1 h. RNA was precipitated into phenol:chloroform (see section 2.2.3). Agarose gel electrophoresis was done to control RNA integrity (see section 2.2.5). Reverse transcription of RNA was carried out as described in section 2.2.4.

Amplification of 5'ends of *A. strigosa* UGT transcripts

Race PCR mix:

Volume	Mix component	concentrations
1 µL	cDNA with 5' end adapters	-
1 µL	dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP)	200 µM each
3 µL	GeneRacer 5' primer (10 µM) (see section 2.1.3)	0.6 µM
1 µL	Reverse gene specific primer (10 µM) (see section 2.1.3)	0.2 µM
5 µL	10x ELT buffer 3	1x
0.5 µL	Expand Long Template DNA polymerase (5 U/µl, Roche)	2.5 U
38.5 µL	steril Milli-Q water	-

Touch down PCR program was done with GS1 thermocycler (G-Storm) as follows:

PCR steps	Duration	Temperature	Cycles
Denaturation	2 min	94°C	x1
Denaturation	30 sec	94°C	x5
Annealing/elongation	2 min	70°C (-1°C/cycle)	
Denaturation	30 sec	94°C	x25
Annealing	30 sec	65°C	
Elongation	1.5min	68°C	
Elongation	10 min	68°C	x1

*annealing temperature is variable depending on primers used (confer primer table)

Nested PCR was done using initial PCR reaction products as templates to increase amplification specificity. PCR mix was carried out as follows:

Volume	Mix component	concentrations
1 µL	Initial PCR product	-
1 µL	dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP)	200 µM each
3 µL	GeneRacer 5′ Nested (10 µM) (see section 2.1.3)	0.6 µM
1 µL	Nested reverse gene-specific primer (10 µM) (see section 2.1.3)	0.2 µM
5 µL	10x ELT buffer 3	1x
0.5 µL	Expand Long Template DNA polymerase (5 U/µl, Roche)	2.5 U
38.5 µL	steril Milli-Q water	-

Nested PCR program was performed with GS1 thermocycler (G-Storm) as follows:

PCR steps	Duration	Temperature	Cycles
Denaturation	2 min	94°C	x1
Denaturation	30 sec	94°C	
Annealing	30 sec	Var*	x25
Elongation	1.5min	68°C	
Elongation	10 min	68°C	x1

*annealing temperature is variable depending on primers used (confer primer table)

Nested PCR products were analysed by gel agarose electrophoresis (see section 2.2.5). DNA fragments of the expected size were purified from an agarose gel slice (QIAquick Gel Extraction Kit, Qiagen).

Purified DNA fragments were cloned into pCR4-TOPO plasmid (see section 2.1.4) using TOPO TA Cloning kit for sequencing (Invitrogen). TOPO cloning reaction was performed according to manufacturer instructions. 6µL of the reaction product were mixed with 50 µL of thawed *E.coli* DH5α cells for heat shock transformation (see section 2.2.10). *E.coli* transformants were selected on LB-G agar supplemented with 50 µg/mL of ampicillin and 10 colonies per gene were used for colony PCR (see section 2.2.12) with M13 primers (see section 2.1.3). Plasmids from colonies with inserts having expected length were purified (see section 2.2.11) and process for sequencing with M13 primers (see section 2.2.6).

2.2.9   Cloning UGT transcripts using a gateway strategy

Amplification of UGT transcripts

A two-step PCR protocol was used, involving an initial amplification with specific primers followed by a second amplification using attB adapters.

UGT transcripts were amplified from *A. strigosa* root tip cDNA templates using primers specific for UGT sequences containing 12 nucleotides of the attB sites at their 5' end (see section 2.1.3). The PCR mix was prepared as follows:

Volumes	Mix component	Final concentration
0.5 μl	cDNA template (100 ng/μl)	50 ng
4 μl	5x Phusion HF buffer	1x
0.4 μl	dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP)	200 μM each
1 μl	FgwUGT specific primer (10 μM)	0.5 μM
1 μl	FgwUGT specific primer (10 μM)	0.5 μM
0.2 μl	DNA polymerase Phusion (2 U/μl)	0.02 U/μl
12.9 μl	sterile Milli-Q water	-

PCR program was performed as follows:

PCR steps	Duration	Temperature	Cycles
Denaturation	2 min	98°C	x1
Denaturation	10 sec	98°C	
Annealing	30 sec	Var*	x10
Elongation	1 min	72°C	
Elongation	1 min	72°C	x1

*annealing temperature is variable depending on primers used (see section 2.1.3).

Products from first PCR step were amplified using attB adapters (see section 2.1.3). PCR mix was prepared as follows:

Volumes	Mix component	Final concentration
10 μl	Product from first PCR	-
8 μl	5x Phusion HF buffer	1x
0.8 μl	dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP)	200 μM each
2 μl	Forward attB1	0.5 μM
2 μl	Reverse attB2	0.5 μM
0.4 μl	DNA polymerase Phusion (2 U/μl)	0.02 U/μl
27 μl	sterile Milli-Q water	-

PCR program was performed as follows:

PCR steps	Duration	Temperature	Cycles
Denaturation	2 min	98°C	x1
Denaturation	10 sec	98°C	
Annealing	30 sec	45°C	x5
Elongation	1 min	72°C	
Denaturation	10 sec	98°C	
Annealing	30 sec	55°C	x20
Elongation	1 min	72°C	
Elongation	7 min	72°C	x1

PCR products were analysed by electrophoresis (see section 2.2.5) and length of DNA fragments was estimated by comparison with 1 kb DNA ladder (New England BioLabs).

Cloning PCR products into pDONR vector

Samples containing a single amplification product with the expected length were purified before the BP clonase reaction. PCR products were diluted 4 times into TRIS-EDTA buffer (TE buffer: 10 mM TRIS-HCl pH 7.5, 1 mM EDTA), ½ volume of polyethylene glycol (PEG) solution (30% PEG 8000, MgCl2 30 mM) was added to the sample. DNA fragments >300 pb were precipitated by centrifugation (15 min, 14,000 g).

Purified PCR products were cloned into a Gateway Donor vector. Reaction mix was prepared as follows:

Volumes	Mix component	Final concentration
x µl	Purified attB PCR product	50 ng
1 µl	pDONR207 (Invitrogen) (100 ng/µl)	100 ng
2.5 µl	TE buffer (10mM)	5 mM
1 µl	BP clonase II (Invitrogen)	-

Samples were mixed by vortexing and incubated for 12 hours at room temperature. The BP clonase protein was inactivated by proteinase K treatment for 10 min at 37°C. Competent DH5α cells were transformed with 1 µl of the reaction product (see section 2.2.10). Transformed cells were selected onto LB-G agar supplemented with 25 µg/mL of gentamicin (Sigma-Aldrich). Presence of the insert into resistant bacteria was confirmed by PCR (see section 2.2.12). Bacterial colonies with the expected DNA fragment were grown for 12 h in 10 mL LB-G supplemented with 25 µg/mL of gentamicin. Plasmid purification was performed on harvested bacterial cultures (see section 2.2.11). Purified plasmids were sequenced (see section 2.2.6) to validate insert identity and avoid mutations.

Transfer of the insert into expression vectors

UGT transcript sequences were transferred from pDONR to expression vectors via LR clonase reaction. Reaction mix was prepared as follows:

Volumes	Mix component	Final concentration
x µl	Purified pDONR207-UGT plasmid	50 ng
1 µl	pH9-GW (100 ng/µL), pEAQ-*HT*-DEST1/2 (150 ng/µL)	100 ng, 150 ng
2.5 µl	TE buffer (10mM)	5 mM
1 µl	LR clonase II (Invitrogen)	-

Samples were mixed by vortexing and incubated for 12 hours at room temperature. Clonase protein was inactivated by proteinase K treatment for 10 min at 37°C. Competent DH5α cells were transformed with 1 µl of the reaction product (see section 2.2.10). Transformed cells were selected on LB-G agar supplemented with 50 µg/mL of kanamycin (Sigma-Aldrich). Presence of the insert into resistant bacteria was monitored by PCR (see section 2.2.12). Bacterial colonies with the expected DNA fragment were grown for 12 h in 10 mL LB-G supplemented with 50 µg/mL of kanamycin. Plasmid purification was performed on harvested bacterial cultures (see section 2.2.11). Purified pH9-GW-UGT plasmids were cloned into *E.coli* BL21 (see section 2.2.10) then selected on LB-G agar supplemented with 50 µg/mL of kanamycin. Purified pEAQ-*HT*-DEST1-UGT and pEAQ-*HT*-DEST2-UGT

plasmids were cloned into *A. tumefasciens* LBA4404 (see section 2.2.10), then selected to LB-G agar supplemented with 50 µg/mL of kanamycin, 50 µg/mL of rifampicin and 100 µg/mL of streptomycin. A selected colony was grown overnight in 10 mL selective LB-G media, then 20% glycerol stocks were made, votexed and frozen into liquid nitrogen. Glycerol stocks were stored at -80°C.

2.2.10 Cell transformation

***E. coli* DH5α/BL21**

Competent cells (10 µL aliquot) were thawed on ice and mixed with 1 µL of clonase reaction product or 10 ng purified plasmid. Samples were incubated 30 min on ice then heat-shocked for 45 sec at 42°C. Samples were placed on ice for 2 min before dilution in 90 µL of S.O.C medium. Cell culture was incubated for 1 h at 37°C, 200 rpm. 50µl of the culture was spread onto LB-G agar supplemented with 50 µg/mL of kanamycin. Solid cultures were incubated over night at 37°C.

***Agrobacterium tumefasciens* LBA4404**

An aliquot of 20 µL of *A. tumefasciens* LBA4404 competent cells was carefully thawed at 4°C for approximately 1 h. The competent cells were mixed with 10µl of pEAQ-HT-DEST1/2-UGT (10ng/µL). The sample was placed into liquid nitrogen for 5min, and then thawed at room temperature. After addition of 1ml LB medium, the sample was incubated for 4 h at 28°C. The culture was spread onto LB-G agar supplemented with 50 µg/mL of kanamycin, 50 µg/mL of rifampicin and 100 µg/mL of streptomycin. Solid cultures were incubated for 2-3 days at 28°C.

2.2.11 Plasmid extraction and purification

Plasmid extractions and purifications were performed using QIAprep Spin Miniprep Kit (Qiagen) following supplier instructions. DNA concentrations were measured by Nanodrop (NanoDrop 8000 spectrophotometer, Thermo Scientific).

2.2.12 Colony PCR

Single colonies were diluted into respective 50 µL aliquots of sterile H$_2$O and heated for 20 min at 95°C. Heated cell samples were centrifuged for 5 min at 14,000 g. The supernatant from each sample was used as template in PCR reactions. PCR mix was prepared as follows:

Volumes	Mix component	Final concentration
2 µl	Heated colony	-
4 µl	GoTaq 5x Green buffer (Promega)	1x
0.5 µl	dNTP mix (10 mM each dATP, dCTP, dGTP, dTTP)	250 µM each
2 µl	MgCl$_2$ (25 mM)	2.5 mM
1 µl	Forward primer (10 µM)	0.5 µM
1 µl	Reverse primer (10 µM)	0.5 µM
0.2 µl	GoTaq DNA polymerase (5 U/µl, Promega)	1 U
9.5 µl	sterile Milli-Q water	-

PCR program was performed as follows:

PCR steps	Duration	Temperature	Cycles
Denaturation	2 min	95°C	x1
Denaturation	30 sec	95°C	
Annealing	30 sec	55°C	x30
Elongation	2 min*	72°C	
Elongation	10 min	72°C	x1

* All UGT transcripts are approximately 1.5kb (except UGT80s)

PCR products were directly analysed by electrophoresis (see section 2.2.5).

2.2.13 Purification of recombinant enzymes

Cell culture and protein production

Transformed *E. coli* cells were grown overnight in 10 mL liquid LB-G plus appropriate antibiotic at 37°C under agitation 300rpm. The whole pre-culture was used to inoculate 5% glycerol liquid LB-G or liquid LB-G only plus appropriate antibiotic (table 4.1). The culture was grown at 37°C, 300 rpm until OD_{600nm} reached 0.6-0.8, and then the temperature was set to 18°C or 21°C at 200 rpm. After acclimation for one hour, expression was induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) (Sigma-Aldrich). After five hours or overnight expression the culture was harvested by centrifugation (7000g, 15min, 4°C). The supernatant was discarded and the pellet was frozen in liquid nitrogen then stored at -80°C until cell lysis.

Cell resuspension and lysis

The frozen pellet was resuspended in 30ml of lysis buffer (300mM NaCl, 50mM TRIS-HCl pH 7.8, 20mM Imidazol, 5% Glycerol, 1mg/ml lysozyme, protease inhibitor (EDTA free, Roche), 10mM β-mercaptoethanol, 0.1% Tween 20 (Sigma-Aldrich)) and shaken (Spiramix 5, Denley) at room temperature for 30 min. Digestion of DNA was achieved by addition of 250 µL of benzonase solution (850 mM $MgCl_2$, 3.8 u/µl benzonase (Novagen)) and incubation at room temperature for 15 min. The resulting cell lysate was centrifuged at 13,000 rpm for 15 min at 4°C. The soluble fraction was filtered before purification (Minisart 0.2 µm, Sartorius).

Purification on HiTrap chelating HP 1mL

Purification of the recombinant enzyme was performed by liquid chromatography (ÄKTA purifier system, GE Healthcare) using an ion exchange column (1 mL HiTrap Chelating HP, GE Healthcare). The soluble fraction from the cell lysate was loaded onto a precooled column charged with Ni^{2+} solution then equilibrated with buffer A (300mM NaCl, 50mM TRIS-HCl pH 7.8, 20mM Imidazol, 5% Glycerol). After complete elution of the unbound fraction, elution of His-tagged recombinant protein was done with a linear gradient of imidazole generated by supplying of buffer B (300mM NaCl, 50mM TRIS-HCl pH 7.8, 500mM Imidazol, 5% Glycerol). The gradient program was as follows: flow rate 0.5

mL/min; isocratic from 0 to 5 min, 0% buffer B; linear gradient from 5 to 15 min, 0-65% buffer B; linear gradient from 15 to 17 min, 65-100% buffer B; isocratic from 17 to 22 min, 100% buffer B; linear gradient from 22 to 23 min, 100-0% buffer B; isocratic from 23 to 28 min, 0% buffer B. Protein elution was monitored with a UV detector at 280nm.

Concentration, storage of the purified recombinant protein

Co-eluted fractions were pooled together and concentrated up to ten times using a concentrator device (Vivaspin 2 mL 10,000 MWCO PES, VIVA science). The purity of the concentrated sample was assessed by SDS-PAGE gel (see section 2.2.18). Protein concentration was determined by the Bradford method (see section 2.2.17). Protein samples were split in 10 µL aliquots at 4°C then directly frozen in liquid nitrogen. Frozen purified protein samples were stored at -80°C.

2.2.14 Enzymatic assays

Radioassays

Reaction mixtures were composed of 100 mM TRIS-HCl pH 7.5, 50 µM UDP-Glc (592 Bq/50 µL) (UDP-Glucose [1-3H], American Radiolabelled Chemicals), 40 µg/mL of recombinant UGT. The reaction was started by addition of 48 µL of reaction mix onto 2 µL of 5 mM acceptor diluted in ethanol. Samples were incubated 12 hours at 35°C, 900 rpm. Samples were transferred onto a 500 µL anion exchange resin (QAE sephadex A-25, Sigma-Aldrich) soaked in water. Reaction products were incubated for 5 min at room temperature with the resin; the tubes were inverted 3-4 times. Samples were diluted with 1 mL of water. Tubes were shortly centrifuged and 800 µL of the supernatant were transferred to a scintillation vial. An additional wash with 800 µL of H_2O was done following the same protocol. The resin was eluted twice with 650 µL of 1M ammonium acetate. The entire remaining volume was transferred to a new scintillation tube with the resin. Radioactive level into unbound (H_2O wash) and bound (ammonium acetate wash) fractions were counted separately with a scintillation counter (Tri-Carb 2910 TR, Perkin Elmer) after addition of 5 mL scintillation fluid (Quicksafe A, Zinsser Analytic) and homogenisation of the solution.

Results presented in section 6.2.1.2 referring to acceptor screening of UGTs were normalised. Two levels of normalisation were used to assess the crude scintillation readings of the bound and unbound fractions in order to take into account the background noise in the assays. Firstly, to remove potential variation in the total amount of radioactivity incorporated in each sample, the percentage of radioactivity in the unbound fraction (products) versus the total radioactivity of the sample (bound and unbound fraction) was estimated. A second level of normalisation was introduced to take into account the background levels of unincorporated radioactivity in each assay. This involved dividing the signal obtained without acceptor from the signal obtained for each acceptor, for each of the different enzymes assayed. Ultimately these values reflect the percentage of signal over the background noise.
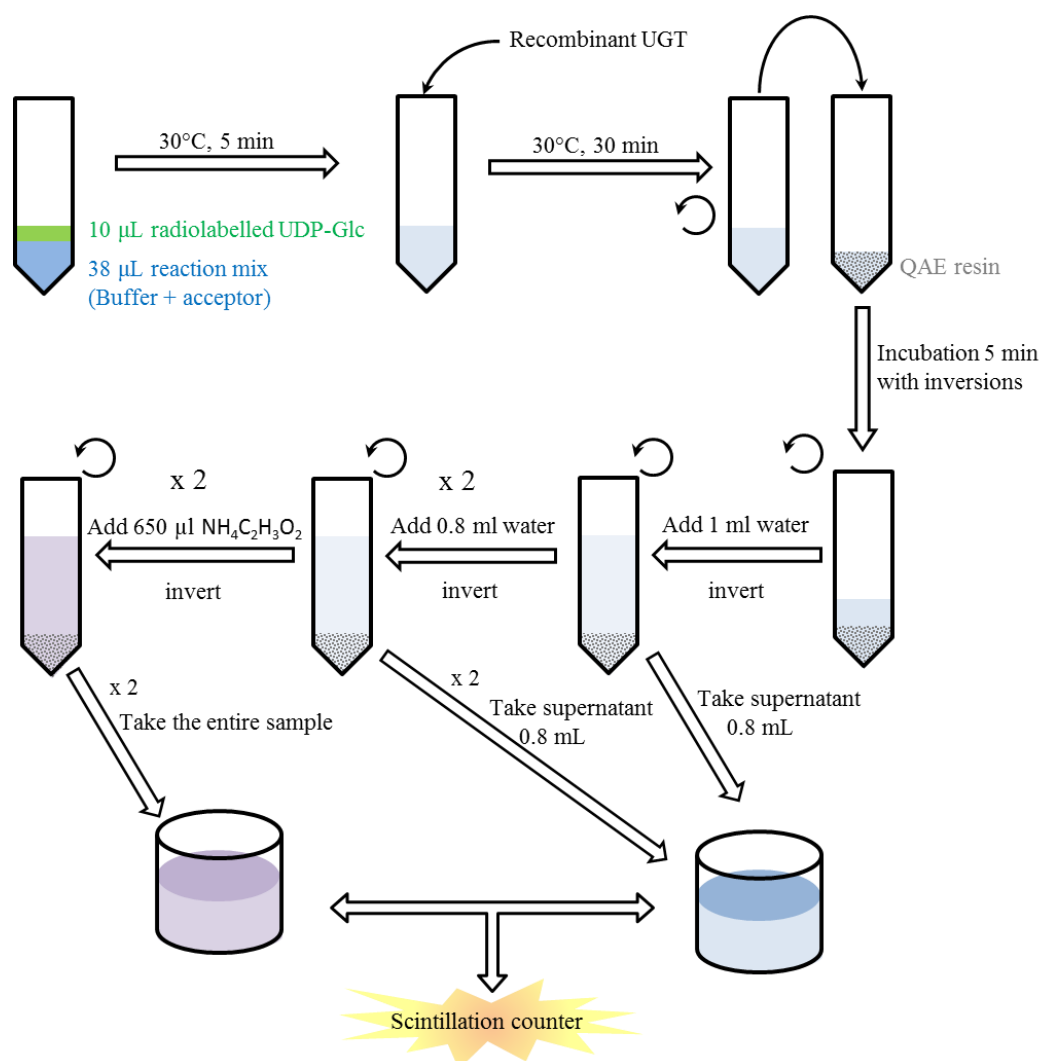
Figure 2.1:   Flowchart showing the layout of glycosyltransferase radioassay. The following sign: ↻ is used for short centrifugation steps.

Glycosylation assay of trichlorophenol

Reaction was made in 50 µl volume, the mix was composed of 100 mM TRIS-HCl pH 7.5, 500 µM of 2,4,5-trichlorophenol (TCP) and 1 mM uridine diphospho sugars (UDP-β-D-glucose, UDP-β-D-galactose or UDP-α-L-arabinose). Reaction was started with the addition of 2 µg of recombinant purified enzyme onto pre-warmed sample. Reaction mix was incubated for 1 hour at 30°C, 300 rpm. Reaction was stopped by the addition of 2 µl of 240 mg/ml trichloroacetic acid. Proteins were precipitated by centrifugation at 14,000 g for 5 min. Supernatant was kept at -20°C until analysis.

Analysis of reaction products was done by reverse phase HPLC coupled with UV detector. The protocol used was derived from Messner et al. (2003). The HPLC system UltiMate 300 LC (Dionex) was used in combination with C18 column (Luna 5 µm C18(2) 100 Å, 30 x 2 mm, Phenomenex). Buffer A was H_2O with 0.1% H_3PO_4; buffer B was 100% acetonitrile. The gradient program was as follows: flow rate 1 mL/min; isocratic from 0 to 2 min, 8% buffer B; linear gradient from 2 to 16 min 8-100% B, isocratic from 16 to 18 min, 100% buffer B; linear gradient from 18 to 20 min 100-8% B; isocratic from 20 to 22 min, 8% buffer B. TCP and TCP glycosides were detected at 205 nm wavelength.

Glycosylation assay of Triterpenoids

Reactions were performed in 100 µl total volume. Each reaction mix was made in 10% ethanol and composed of 100 mM TRIS-HCl pH 7.5, 200 µM of triterpenoids, 1 mM uridine diphospho- sugars (UDP-β-D-glucose or UDP-α-L-arabinose). Reactions were started with the addition of 2 µg of recombinant purified enzyme in pre-warmed samples. Reaction mixtures were incubated for 12 hours at 35-37°C, 900 rpm. Reactions were stopped and products were extracted twice in 500µL ethyl acetate after dilution into 400 µL Milli-Q H_2O. The organic fraction of each reaction was entirely dried under nitrogen flux and resuspended into 50 µL methanol.

For Thin Layer chromatography (TLC), plates were loaded with 20 µL of the methanolic extracts from samples as described above. Samples were pre-run 3 times in 100% methanol 0.5 cm above the loading line. The mobile phase was composed of dichloromethane:methanol:H_2O (80:19:1; v:v:v). Plates were sprayed with

methanol: Sulfuric acid (9:1; v:v) then heated at 130°C for 2-3 min. Pictures of stained plates were taken under UV transilluminator.

2.2.15 NMR analysis

Proton NMR spectra were recorded in deuterated water on a Bruker Avance III spectrometer at 400 MHz. Chemical shifts are reported with respect to residual hydrogen deuterated oxide at δH 4.70 ppm, if not stated otherwise.

2.2.16 Protein extraction

Desired plant tissue was collected and grinded in liquid nitrogen (100 mg). The powdered tissue obtained was resuspended in 500 mL of protein extraction buffer (50 mM Tris-HCl pH 7, 150mM NaCl, 10% glycerol, 1% polyvinylpolypyrrolidone (PVPP, Sigma-Aldrich),1% triton x100 (Sigma-Aldrich), 5µM DTT, Protease inhibitors (EDTA-free tablets, Roche)). The tissue was disrupted using pellet pestles then the sample was shaken at 300 rpm for 1 hour at 4°C. Residual tissue particles were removed by centrifugation (14,000 rpm, 30 min at 4°C). Protein concentration was estimated by Bradford method (see section 2.2.17).

2.2.17 Protein quantification by Bradford method

Multiple dilutions of Bovine Serum Albumin (BSA $\geq$ 96%, Sigma-Aldrich) were prepared (2, 1, 0.75, 0.5 and 0.25 mg/mL). Protein samples were diluted up to 20 times and 10 µL of diluted protein samples and BSA dilutions were mixed with 790 µL of H_2O and 200 µL of Bradford reagent (Protein assay, Bio-Rad). Samples were left at room temperature for 15-30 min. Absorption of the samples was monitored at 495nm (UVmini 1240, Schimadzu) after blanking with 800 µL of H_2O mixed with 200 µL of Bradford reagent. Standard linear regression curve was drawn with BSA absorbance values and the equation of the curve was used to estimate protein concentration in tested samples.

2.2.18 <u>Protein electrophoresis</u>

Protein electrophoresis was done with mini-cell system (XCell SureLock Mini-Cell Electrophoresis System, Invitrogen). Protein samples were prepared as follows: 1x LDS loading buffer (NuPAGE LDS Sample Buffer, Invitrogen) and 30-10 µg of protein. Samples were heated 15 min at 95°C. SDS-PAGE gel (NuPAGE 4–12% Bis-Tris, Invitrogen) was loaded with 20 µL of samples (5 µL for the protein ladder (SeeBlue Pre-Stained Standard, Invitrogen)) A current of 200 volts was applied to the gel for 30-50 min. Gel was stained onto coomasie solution (InstantBlue, Expedeon) for 20-30 min.

2.2.19 <u>Western blot</u>

After separation of protein preparations by SDS-PAGE (see above) proteins were transferred to a 0.2 µm nitrocellulose membrane (Bio-Rad) using Novex mini-cell (Invitrogen) according to manufacturer instructions. To control blotting, the membrane was stained by Ponceau S staining solution (Sigma-Aldrich) then washed several times with distilled water until complete disappearance of the signal. The membrane was washed twice in Tris-buffered saline solution (TBS; 150 mM NaCl, 10mM TRIS pH7.5) before blocking in TBS buffer with 5% powdered milk for 1 hour. The membrane was washed twice in TBS Tween/Triton (TBSTT; 500 mM NaCl, 20mM TRIS pH7.5, Tween 20 (Sigma-Aldrich), Triton X-100 (Sigma-Aldrich)) then once in TBS for 10 minutes. The membrane was incubated for 1 hour with the primary antibody (Monoclonal Anti-polyHistidine, Sigma-Aldrich) diluted 1:1000 in blocking solution. The membrane was washed twice in TBSTT then once in TBS. The membrane was incubated for 1 hour with the secondary antibody (Anti-Mouse IgG Alkaline Phosphatase Conjugate, Sigma-Aldrich) diluted 1:5000 in blocking solution. The membrane was washed five times in TBSTT before staining with 5 mL of nitro-blue tetrazolium chloride (NBT) / 5-bromo-4-chloro-3-indolyl phosphate (BCIP) solution (100 mM TRIS-HCl pH9, 150 mM NaCl, 1mM MgCl2, NBT 330mg/ml (Promega), BCIP 165 mg/ml (Promega)) for 1 hour in the dark.

2.2.20 <u>Proteomic analysis (from G.Saalbach)</u>

The gel slice was washed, reduced and alkylated, and treated with trypsin according to standard procedures adapted from Shevchenko et al. (1996). Peptides were extracted with 5% (v/v) formic acid/50% acetonitrile (v/v), dried down, and re-dissolved in 0.1% (v/v) TFA. Nano-LC-MSMS experiments were performed on an LTQ-OrbitrapTM mass spectrometer (Thermo Fisher Scientific Inc., Hemel Hempstead, UK) coupled to an EASY-nLC HPLC via an ion source (Proxeon, Odense, Denmark). Aliquots of the extracted peptides were loaded onto a C18 PepMapTM trap column (Dionex, Camberley, UK) which was then switched in-line to an analytical column (BEH C18, 1.7 μm, Waters, 75 μm x 120 mm, self-packed) for separation. The LC system was run at a flow rate of 250 nL/min with a gradient of 5-40% acetonitrile in water/0.1%formic acid at a rate of 1% min^{-1}.

The mass spectrometer was operated in positive ion mode. Raw files were processed with MaxQuant version 1.3.0.5 (Cox and Mann, 2008; http://maxquant.org) to generate re-calibrated peaklist-files which were used for a database search using an in-house Mascot® 2.4 Server (Matrix Science Limited, London, UK). Mascot searches were performed on a custom database containing all available UGT protein sequences from *A. strigosa* root tip in a background of 1000 random *A. thaliana* sequences downloaded from Uniprot (<u>www.uniprot.org</u>). Mascot search results were imported and evaluated in Scaffold 3.6.3 (proteomsoftware.com, Portland, OR, USA).

2.2.21 <u>Phylogeny</u>

The Carbohydrate-Active enZymes, CAZy database (http://www.cazy.org/) was used to collect amino acids sequences from characterised plant GT1s. NCBI protein BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi) was used in parallel to find homologues of *A. strigosa* GT1s. A selected set of sequences was aligned using MAFFT (http://www.imtech.res.in/raghava/mafft/). The unrooted phylogenetic tree was constructed using MEGA 4 (http://www.megasoftware.net/) by the neighbor-joining method.

2.2.22 Transient expression *in Nicotiana benthamiana*

Transformed *A. tumefaciens* cells were grown under antibiotic selection (rifampicin + streptomycin + kanamycin) on LB-agar plates at 28°C for 3 days. Each isolated colony was inoculated in 50 mL liquid LB medium under antibiotic selection and grown for 24h at 28°C. Bacterial cultures were harvested by centrifugation at 4000 g for 15 min at 4°C. The resulting pellet was resuspended in 5 mL of agromix (10 mM MES-KOH, 10 mM MgCl2, 150 μM acetosyringone) then incubated at room temperature in a dark place for 2h. Optical density of the culture was measured at 600nm (UV-Vis Spectrophotometer, Shimadzu). The culture was diluted down to 0.8 OD_{600nm}. For co-infiltration the same volume of each bacterial dilution was mixed together. Sub-epidermal infiltration of the diluted agromix was performed into *N.benthamiana* leaves of 3 weeks old plants. Six days after infiltration, the infiltrated area of the leaves were collected and directly frozen in liquid nitrogen. Leaf tissue was stored at -80°C freezer until use.

2.2.23 Analysis of metabolites from *Nicotiana benthamiana* leaves

Frozen *N.benthamiana* leaves were ground with pre-cooled mortar and pestle. Extraction was performed with 100 mg of ground tissue into 1 mL 75 or 40% methanol for 12h under agitation at 4°C. Samples were filtered (minisart 0.2 μm, Sartorius) and freeze-dried. Dried material was resuspended into 30 μL methanol and the whole solution was loaded onto TLC plate. The TLC plate was pre-run 3 times in 100% methanol 0.5 cm above the loading line. The mobile phase was composed as follows dichloromethane:methanol:H_2O (80:19:1; v:v:v). Plates were sprayed with acetic acid:Sulfuric acid:*p*-anisaldehyde (48:1:1; v:v:v) then heated at 130°C for 2-3 min until coloration appeared.

2.2.24 Purification of 12,13-epoxy-16-hydroxy-β-amyrin-3-*O*-glucoside from *N. benthamiana* leaves

Grinded material was extracted at 35°C, 300 rpm for 1h in 40% methanol. The soluble fraction was recovered after centrifugation at 15,000g for 20 min and filtered (0.2 μm, minisart filters). The methanolic extract is then partitioned three times in hexane. The obtained aqueous phase is then partitioned three times in ethyl acetate. The resulting organic fraction was dried and resuspended in toluene for further separation by MPLC (Medium Pressure Liquid Chromatography) using the Biotage system (Biotage, Uppsala, Sweden). The material was loaded on a pre-packed silica gel column (12-S column, Biotage) and eluted by running a linear gradient of dichloromethane:methanol (A:B) with a flow rate of 6 mL/min; 0-6 min: 0% B; 6-48 min: 0-100% B; 48-60 min: 100% B. Fractions (6mL) were collected and assayed by TLC using dichloromethane:methanol:H_2O (80:19:1; v:v:v) as solvent; components were detected with p-anisaldehydey/sulfuric acid/acetic acid (1:1:48, v/v/v). All fractions containing the product were combined, dried and resuspended in 50% methanol for further purification using semi-preparative HPLC. The sample was applied on C18 semi-preparative HPLC column (Phenomenex, Luna 5μ C18(2) 100A, 250 x 10 mm, 5 micron) coupled with Charged Aerosol detector (CAD, Corona Ultra RS from Dionex). Chromatography was performed with a flow rate of 3 mL/min. solvent A: H2O, solvent B: CH3CN. A linear gradient was used: 0-5 min: 0% B; 5-30 min: 0-100% B; 30-35 min: 100% B; 35-36 min: 100-0% B; 36-41 min 0% B. The fractions containing the analyte (by TLC) were collected, pooled and evaporated to dryness.

2.2.25 LC-MS analysis

Method used in triterpenoid glycoside analysis

Sample were run on a Surveyor hplc equipped with a DecaXPplus ion trap MS (Thermo). Analytes were separated on a 100×2mm 3μ Luna C18(2) (Phenomenex) column using a gradient of 0.1% formic acid in water (buffer A) versus acetonitrile (buffer B), run at 300 μL.min-1 and 30ºC. The gradient program was as follows: linear gradient from 0 to 3 min 20-25% B, linear gradient from 3 to 20 min 25-60% B, linear gradient from 20 to 30 min 60-90% B, isocratic from 30 to

32 min, 90% buffer B; linear gradient from 32 to 33 min 90-20% B; isocratic from 33 to 45 min, 20% buffer B. MS data was collected by electrospray from m/z 150-1500, in positive and negative modes (in separate runs), with 50 units sheath gas, 5 units aux gas, 350ºC capillary temperature, and 5.2kV spray voltage (positive) or 5.0kV (negative). Full MS was collected and data dependent MS2 and MS3 data at 35% collision energy and an isolation width of 3.0 amu. MS2 was at an isolation width of m/z 4.0 and 35% collision energy. Spectrometric Detection was by UV/vis absorbance collecting spectra from 200-600nm.

Method used in flavonoid glycoside analysis

Analytes were separated using a different HPLC program and column; separation with 100×2mm 3μ Luna C18(2) column (Phenomenex) using a gradient of 0.1% formic acid in water (buffer A) versus acetonitrile (buffer B), run at 300 μL.min-1 and 30ºC. The gradient program was as follows: linear gradient from 0 to 15 min 5-95% B, isocratic from 15 to 17 min, 95% buffer B; linear gradient from 17 to 17.5 min 95-5% B; isocratic from 17.5 to 21 min, 5% buffer B. Detection was done by UV/visible absorbance collecting spectra from 200-600nm as well as electrospray MS.

Method used in trichlorophenol glycoside analysis

Analytes were separated using a different HPLC program and column; separation with 50×2.1mm 2.6μ Kinetex XB-C18 column (Phenomenex) using a gradient of 0.1% formic acid in water (buffer A) versus acetonitrile (buffer B), run at 300 μL.min-1 and 30ºC. The gradient program was as follows: linear gradient from 0 to 10 min 25-60% B, linear gradient from 10 to 20 min 60-95% B, isocratic from 20 to 23 min, 95% buffer B; linear gradient from 23 to 24 min 95-25% B; isocratic from 24 to 29min, 25% buffer B. Detection was done by UV/visible absorbance collecting spectra from 200-600nm as well as electrospray MS.

Chapter 3 - Identification of root-expressed *A. strigosa* UGTs

3.1 Introduction

The family one glycosyltransferases (UGTs) are involved in various essential processes in plants, including cellular homeostasis, defence and detoxification of xenobiotics (Bowles et al. 2005). Considerable progress has been made in functional characterization of UGTs from dicots, most notably in the model plant *A. thaliana* where systematic approaches have been taken to investigate the full UGT complement encoded by the *A. thaliana* genome (Caputi et al. 2008; Hou et al. 2004; Li et al. 2001; Lim et al. 2003; Lim et al. 2002; Weis et al. 2006). So far, very little is known about the function and evolution of UGTs in monocots despite their potential agronomical significance. Cereals are essential to human and livestock nutrition and many glycosides are known to be involved in defence and adaptation to environmental conditions (Fay and Duke 1977; Macias et al. 2006; Maier et al. 1995).

Roots accumulate a wide variety of specialised metabolites with important roles in defence against soil-borne pathogens, allelopathy, interaction with symbiotic organisms, nutrient uptake and also phytohormone-mediated signalling (D'Auria and Gershenzon 2005; Flores et al. 1999; Weston and Mathesius 2013). Glycosides represent a large proportion of the specialised metabolites found in roots. In cereals, the roots of rice have been reported to produce flavone *O*-glycosides as potential allelochemicals (Kong et al. 2007) and benzoxazinoids are found in the roots and other organs of the Poaceae where they are stored in glycosylated forms (Macias et al. 2006). Apart from seeds, little is known about the metabolite composition of *A. strigosa* tissues. Like other cereals (Brazier-Hicks et al. 2009), oats are known to synthesise flavone-*C*-glycosides that are implicated in protection against parasitic nematodes (Soriano et al. 2004) and avenacins, which protect oats against take-all disease (Papadopoulou et al. 1999). A comprehensive and systematic analysis of UGTs expressed in *A. strigosa* roots will therefore be an important step in investigating the glycosylation of small molecules in cereals, including those involved in plant defence.

3.1.1   Phylogeny of plant family one glycosyltransferases

As mentioned previously (see Chapter 1), the classification of family one glycosyltransferases (or UGTs) is based on amino acid sequence similarity (Mackenzie et al. 1997). The plant UGTs described so far have been classified into families, from UGT70 to UGT100; those families have been split in 16 monophyletic groups from A to P as described in section 1.2.5 and figure 1.3. (Caputi et al. 2011; Ross et al. 2001).

Recent genome-wide approaches exploiting genome sequences from multiple plant species have provided insights into the evolution of UGTs in plants. Like other multigene families implicated in secondary metabolite biosynthesis, UGTs have expanded dramatically in higher plants (Caputi et al. 2011; Yonekura-Sakakibara and Hanada 2011). Extended analyses of UGT families in sequenced species from the plant kingdom reported 12 putative UGT sequences in *Physcomitrella patens*, 74 sequences in *Selaginella moellendorffii* and more than a hundred sequences in most higher plants (Caputi et al. 2011). Large UGT families are also found in the genomes of monocot species, for example 168 UGT genes in *Z. mays*, 201 in *S. bicolor*, 143 in *Brachypodium distachyon* and 213 in *O. sativa* (Yonekura-Sakakibara and Hanada 2011). The lineage-specific evolution of UGTs makes prediction of activity based on sequence similarity alone a tricky task. In the case of oat, prediction is made even more difficult due to the low number of functionally characterised UGTs in monocot species.

3.1.2   Relationship between phylogeny and function of plant UGTs

Over the last decade considerable effort has been invested in the functional characterisation of multiple plant UGTs from various UGT families from a variety of plant species. Reconstruction of the UGT phylogeny helps to make sense of the functional evolution in this superfamily. Early functional studies on flavonoid glycosyltransferases have shed light on the regiospecificity displayed by a majority of UGTs. Three major clades of flavonoid glycosyltransferases have been identified based on UGT phylogeny: Clade I (UGT78s) is composed of flavonoid 3-*O*-glycosyltransferases; Clade II (UGT75) contains flavonoid 5-*O*-glycosyltransferases; and Clade III (UGT73s, UGT88s, UGT89s) is composed of flavonoid 7-*O*-

glycosyltransferases (Noguchi et al. 2009; Tohge et al. 2005; Yamazaki et al. 1999). Therefore, molecular structure of the sugar acceptor appears to have had a relatively low impact on UGT evolution, and a great diversity of compounds can be recognised by a single UGT family. The two divergent families UGT80 and UGT81 are the only families of plant UGTs that glycosylate restricted classes of compounds (sterols and lipids, respectively) (Grille et al. 2010; Jorasch et al. 2000).

Little is known about regiospecific sugar transfer onto triterpenoids, compared to the more extensively studied flavonoids. Nevertheless, members of the UGT73 family appear to be major players in triterpenoid glycosylation (see section 1.3.3). Acceptors that have structural similarities with triterpenes, such as brassinosteroids and steroidal compounds, are also known to be glycosylated by UGT73s (Husar et al. 2011; Itkin et al. 2013; Itkin et al. 2011; Kohara et al. 2005; Kohara et al. 2007; Moehs et al. 1997; Poppenberger et al. 2005).

The identity of the functional group used as the acceptor of the sugar transfer might also play a role in functional evolution of plant UGTs. This is illustrated by the group L mainly composed of ester-forming UGTs, especially in families UGT84 and UGT74 (Owatworakit et al. 2012; Yonekura-Sakakibara and Hanada 2011). UGTs from group A (UGT79s, UGT91s and UGT94s) are also predominantly involved in the transfer of sugars onto a glycosidic moiety of plant natural product glycosides, they were called GGTs by K. Saito and co-workers (Yonekura-Sakakibara et al. 2012). Those enzymes may be key in the formation of sugar chains in the synthesis of plant secondary metabolites such as avenacins.

3.1.3   Identification of components of metabolic pathways through transcript profile analysis

Synthesis of specialised metabolites in plants is generally restricted to a specific tissue or triggered in response to particular biotic/abiotic stresses (Hartmann 2007). For example, flowers synthetize aromatic compounds to attract pollinators and phytoalexins accumulate in tissues that have been challenged by pathogens. Therefore, a number of systems biology approaches to identify the genes of particular specialized metabolic pathways have been based on transcript profile analyses (Schilmiller et al. 2012). Analyses of transcript abundance in various tissues allows candidate genes implicated in the synthesis of particular metabolites to be

identified through comparison of their expression profiles with those of genes that are known to be part of the given metabolite pathway or by virtue of their co-ordinate expression under the conditions where the metabolite of interest is synthesized (Yonekura-Sakakibara et al. 2008).

Co-expression analysis was used successfully for identification of biosynthetic genes for saponins in *Medicago truncatula* (Naoumkina et al. 2010). In *A. strigosa*, avenacin biosynthesis is restricted to the epidermal cell layer of the root tip. All of the cloned and characterised *Sad* genes have been shown to be tightly co-expressed in avenacin accumulating tissues (Fig. 1.8). Consequently a co-expression analysis offers a potential route to uncover new genes and enzymes involved in avenacin biosynthesis.

3.1.4 <u>Aims</u>

To date very little is known about enzymes that catalyse glycosylation of root metabolites in monocots, and oats in particular. The aim of the work presented in this chapter is to collect and analyse a comprehensive collection of UGTs expressed in the root tips of *A. strigosa*. Mining of transcriptomic dataset from 454 sequencing of oat root tip was performed to identify UGT sequences. Phylogenetic analysis of the resulting collection was carried out to classify these UGTs and gain insights into their potential activities. Proteomic analysis was carried out in parallel as a complementary strategy to investigate those UGTs that were present preferentially in *A. strigosa* root tips compared to the older parts of the root. A subgroup of potential triterpene glycosyltransferases was identified. In order to evaluate the possible participation of these candidates in avenacin biosynthesis, the expression profiles of these UGT genes in different tissues were compared with those of characterized *Sad* genes.

3.2   Results and discussion

3.2.1   Screening the oat root tip transcriptome for UGT candidates

3.2.1.1   *Identification of a collection of 26 UGT sequences expressed in A. strigosa root tips*

Prior to the start of this project a collection of more than 16,000 expressed sequence tags (ESTs) had been generated from root tips of diploid oat (*A. strigosa*), as described by Haralampidis et al. (2001). This *A. strigosa* EST collection was mined for UGT sequences. TFASTA searches lead to the identification of 26 putative UGT sequences. Rapid Amplification of cDNA Ends (RACE) PCR was conducted using primers specific to these partial UGT sequences. Full-length sequences were obtained for 19 UGT candidates (Townsend et al, unpublished data). Among those genes, *Ugt74H5* (*Sad10*) was found to be part of the avenacin gene cluster. Biochemical characterisation and gene expression analysis suggest UGT74H5 catalyses the synthesis of the acyl glucose donor used in avenacin biosynthesis (Owatworakit et al. 2012).

Completion of the collection of full-length coding sequences (CDSs) was achieved by RACE PCR. The 5' ends of six of the remaining seven UGT sequences were obtained (GT27f7, GT3i21, GT25n16, GT14h20, GT5d1 and GT8i4). Conditions for the specific amplification of the 5' end of the seventh sequence, AsGT14b16, were not found.

3.2.1.2 *Mining of a new 454-based transcriptomic resource from A. strigosa to augment the UGT collection*

Recent developments in transcriptome pyrosequencing opened up opportunities for deep sequencing of the *A. strigosa* root tip transcriptome. A transcriptomic analysis was performed by TGAC genomics team (Norwich Research Park) using purified mRNA extracted from the terminal 0.5 cm of *A. strigosa* root tips to augment the transcriptome resource established earlier. Massively parallel pyrosequencing was undertaken using 454 titanium chemistery (Roche) to generate in excess of 600 Mpb of sequence reads. Individual reads were processed through Newbler to assemble the individual reads into transcripts. A total of 14,829 contigs were assembled. A BLAST tool was then developed by TGAC bioinformatics team enabling comparison of query sequences.

Transcripts corresponding to previously characterised avenacin biosynthetic *Sad* genes were identified within this contig collection (table 3.1). Each of the five contigs identified contained the full-length coding sequence (CDS) plus flanking 5' and 3' UTR sequences. This demonstrates the usefulness of the pyrosequencing technique for assembly and identification of root-expressed *A. strigosa* genes, including the remaining uncharacterised genes in avenacin biosynthesis.

Table 3.1: Sad gene contigs obtained from 454-based transcriptomic analysis.

Sad gene	Contig number	Number of reads	CDS size (bp)	Contig size (bp)
Sad1	09590	1132	2271	2619
Sad2	17168	918	1470	1801
Sad7	25523	170	1479	1602
Sad9	16631	375	1062	1243
Sad10	07600	71	1392	1667

The 454 contig collection was then mined for UGT sequences. These were identified through BLAST (tBLASTn) searches using the conserved 44 amino acid PSPG motif and also through searches using entire UGT sequences from *A. thaliana* representing each UGT family (Supp. S.1). A total of 110 unique UGT-like sequences were identified (Supp. S.2), 53 of which were predicted to correspond to entire CDSs (Table 3.2).

Six of the 26 UGT sequences found previously (exploiting the ESTs collection) were not represented in the 454 contig collection. Variation in experimental conditions (growth conditions or handling) may explain these differences. Assembly errors can generate a single contig out of two homologous transcripts. Some incorrectly assembled transcripts were identified during the subsequent cloning of the UGT CDSs; contig23586 has been assembled with reads from two homologous transcripts GT23586A and GT23586B (see section 4.2.1). Some assembly errors are inevitable, particularly given the absence of a reference genome for oat.

The number of UGT sequences found in this screen is consistent with the size and expansion of the UGT family in higher plants, and in monocot species in particular. The UGT-like contigs found in the *A. strigosa* root tip transcriptome represent 0.37% of the entire contig collection. This is in accordance with reports that UGT genes represent 0.18 - 0.67% of the total protein-coding genes in the genomes of sequenced higher plants (Yonekura-Sakakibara and Hanada 2011). It is worth to mention that oat genome (11,300 Mb) is much larger than rice genome (430 Mb) and approximately equivalent to the wheat one (17,000bp). Clearly, the analysis described here is based on expressed genes rather than the complete genome sequence, and is restricted to those UGT genes that are expressed in roots. This comprehensive collection of UGT sequences now opens up the opportunity to take a systematic approach towards investigation of UGTs that are expressed in oat roots and, more specifically, to mine for UGTs implicated in avenacin glycosylation. However, working with such a large amount of sequences is challenging and appropriate tools to discriminate these GTs are essential.

3.2.2   Oat UGT phylogenetics

The growing number of functional studies of UGTs have brought new insights into the relationship between amino acid sequences and functions of members of this large enzyme family (Yonekura-Sakakibara and Hanada 2011). Investigation of the phylogenetic relationships of the oat UGTs retrieved from transcriptomic analysis will enable these enzymes to be classified into families and subfamilies and may also yield insights into potential function.

The Carbohydrate Active Enzyme (CAZY) database is a central resource of information for carbohydrate-modifying enzymes (http://www.cazy.org/). A comprehensive collection of functionally characterised plant UGTs was collected from the CAZY website and completed with recent reports on functional characterisations of plant UGTs. A phylogenetic analysis was conducted with these sequences and full-length *A. strigosa* UGT sequences obtained from the transcriptomic analysis (Fig. 3.1). Accession numbers and corresponding literature of UGTs included in phylogenetic tree are available on supplementary data (Supp. S.3). The *A. strigosa* sequences were named following there accession number from the original EST collection or the accession number of the corresponding contig from the 454 sequencing project; those accession numbers were precede by "AsGT" for *A. strigosa* glycosyltransferases. This nomenclature was adopted pending the submission of oat UGT sequences to the international nomenclature committee.
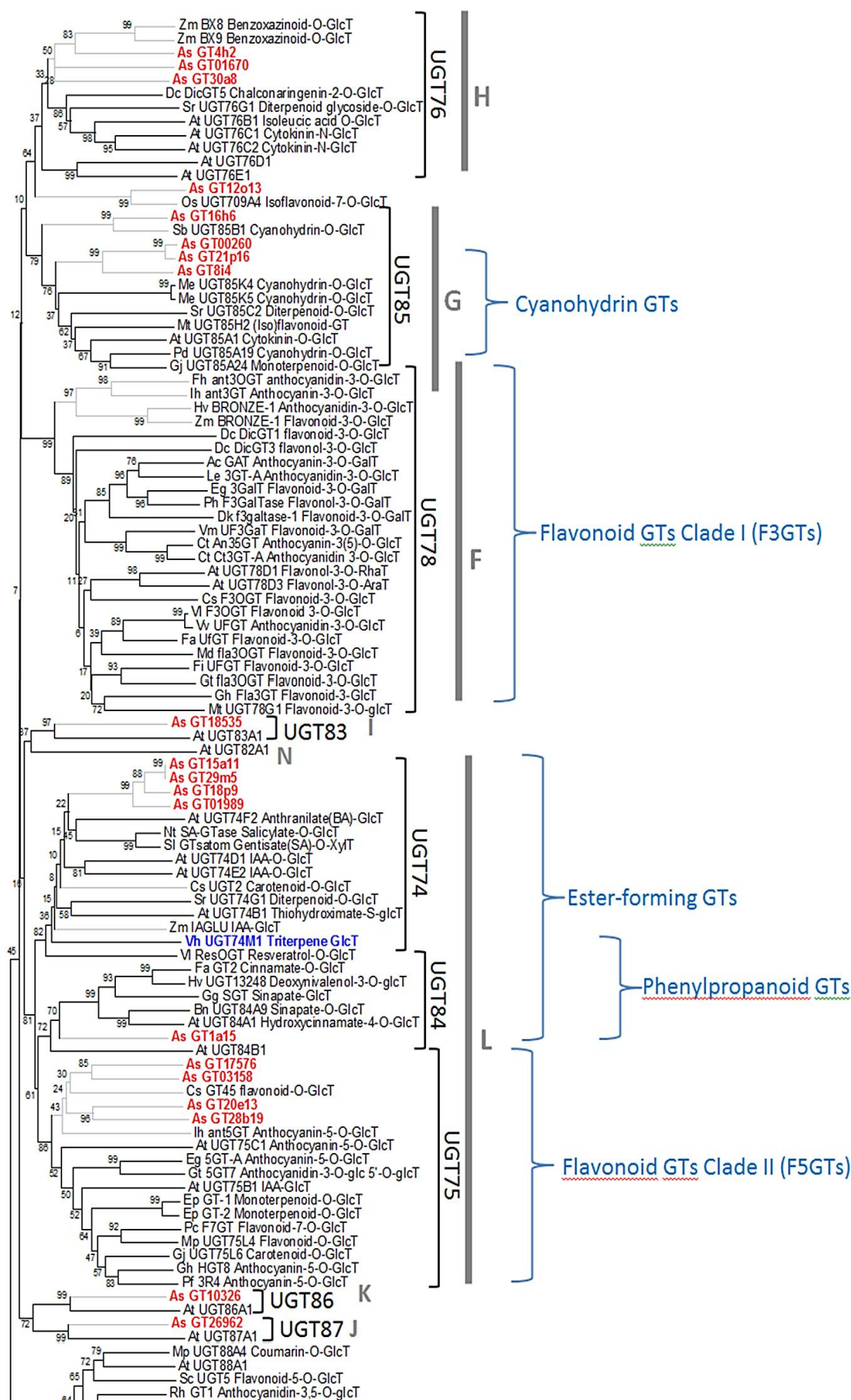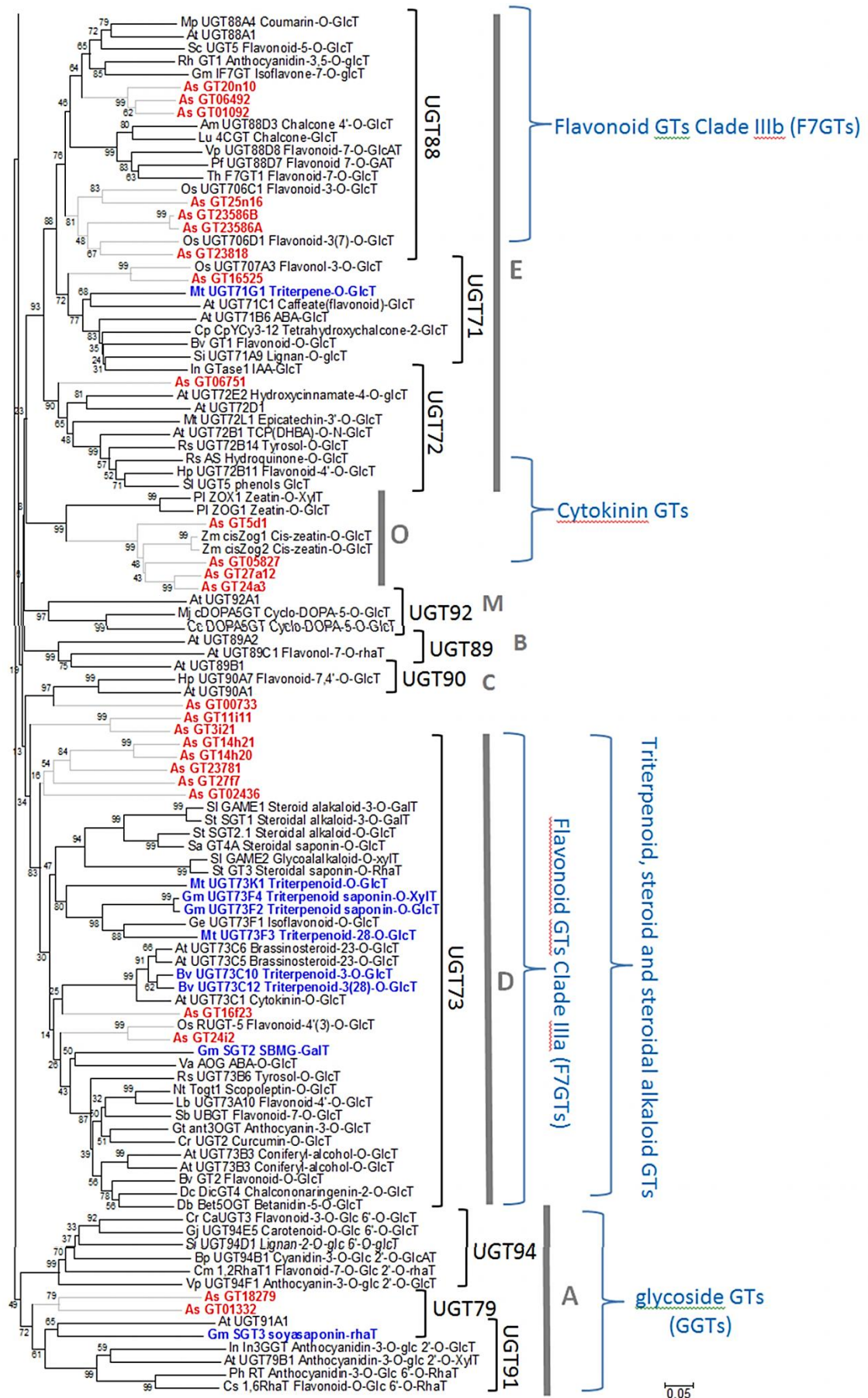
Figure 3.1: Reconstruction of the plant UGT (family 1 glycosyltransferase) phylogeny.
The deduced amino acid sequences from *A. strigosa* (in red) were aligned with those of functionally characterised UGTs form other plant species. Triterpenoids glycosyltransferases are indicated by bold blue characters. Monocot specific branches are drawn in light grey. The phylogenetic tree was drawn using the Neighbor-Joining method. The accuracy of the tree topology was estimated by bootstrap analysis (500 replicates). Accession numbers and references of the sequences used in this phylogenetic tree are given in supplementary data (Supp. S.3).   The scale bar represents 5% divergence.

3.2.2.1   *Oat UGTs are spread through the plant glycosyltransferase phylogeny*

The overall tree topology is in accordance with previous plant UGT phylogeny reconstructions (Caputi et al. 2011; Li et al. 2001). The UGT families and monophyletic groups described by Li et al. (2001) were all identified within the tree. The oat UGTs are distributed across the UGT phylogeny.

The majority of *A. strigosa* UGT sequences are clustered within the plant UGT phylogenetic tree (Fig. 3.1). This is likely to be a reflection of lineage-specific evolution of *A. strigosa* UGTs, in accordance with observations from other plant species (Caputi et al. 2011). *A. strigosa* UGT sequences are localised in monocot-specific branches. Monocot/dicot separation into UGT families has previously been observed for phylogenetic trees of genome-wide collections of UGTs from multiple plant species (Caputi et al. 2011). Interestingly, despite the substantial evolutionary distances between dicot and monocot sequences, activity is generally retained between dicot and monocot enzymes of the same family. This phenomenon is particularly obvious for members of the UGT78 and UGT74 families, where activities from sufficient monocot and dicot enzymes have been described to each group to support this conclusion. This implies that these UGT enzymes have retained the activity of an ancestral enzyme that appeared before the monocot/dicot split.

3.2.2.2   *Repartition of oat UGTs in GT1 families*

The UGT73, UGT85 and UGT88 families constitute the most common families represented in the transcriptomic data from *A. strigosa* root tips (Fig. 3.1), constitute 11%, 11% and 13% respectively of the entire collection of full-length UGTs. A number of UGT73s have previously been shown to glycosylate triterpenes and related compounds in other plant species. The phylogeny of the *A. strigosa* UGT73 family is examined in further details in figure 3.3. Many UGT88s form part of clade IIIb of flavonoid glycosyltransferases, those enzymes are reported as flavonoid-7-*O*-glycosyltransferases (Noguchi et al. 2009; Ono et al. 2010b). The functions of UGT85s enzymes remain to be clarified. Nevertheless, several UGTs from this family catalyse the formation of cyanogenic glucosides (Hansen et al. 2003; Kannangara et al. 2011; Takos et al. 2011).

Substantial clusters of *A. strigosa* sequences are present in group L and group O. The representatives of the UGT75 and UGT74 families from group L each comprise 7% of the total collection. The majority of characterised group L enzymes are ester-forming glycosyltransferases. Group O enzymes are not represented in *A. thaliana.* Consequently functional analysis of this group is limited and the functions of group O enzymes remain to be clarified. Group O enzymes are found in all other sequenced plant (monocot and dicot species) this group is also absent in *Physcomitrella patens* and *Selaginella moellendorffi.*

Other *A. strigosa* sequences are isolated in the tree; many of these belong to small UGT families for which the biological significance is not understood due to a lack of functional analysis. Consequently, we have no indication of the likely activity of these *A. strigosa* UGTs with the exception of AsGT1a15, part of the UGT84 family containing several phenylpropanoid ester-forming glycosyltransferases. AsGT01332 and AsGT18279 (family UGT91) may, like UGT94s and UGT79s, be involved in glycosylation of sugar moieties as suggested by Yonekura-Sakakibara et al. (2012). The β1-2 rhamnosyltransferase activity of UGT91H4 towards glycosidic moiety of soyasaponin precursor (Fig. 1.6) was not mentioned in this publication and reinforces this hypothesis.

AsGT3i21 and AsGT11i11 are related sequences that share 61% amino acid identity. They have less than 40% identity with other characterised sequences and therefore may be part of a new UGT family. They share 53% and 60% identity repectively with the closest *B. distachyon* sequence. Their distant phylogenetic relationship to other UGTs suggests that these enzymes may have evolved new specific functions.

Five UGT families are absent from *A. strigosa* root tip transcriptome (Fig. 3.1): UGT78, UGT79, UGT89, UGT92 and UGT94. The absence of UGT78 sequences in *A. strigosa* is in accordance with similar observations for the sequenced monocot species, *Oryza sativa* and *Sorghum bicolor* (Caputi et al. 2011). Interestingly, a total of four UGT78 monocot sequences have been functionally characterised from other monocots (*Zea mays*, *Hordeum vulgare, Iris hollandica* and *Freesia hybrida*). These form a monocot-specific clade within group F; unfortunately, gene expression analysis was not conducted in roots for any of these enzymes (Ford et al. 1998; Sui et al. 2011; Yoshihara et al. 2005). The other four UGT families are represented in *O. sativa* and *S. bicolor* but have not expanded in these species to the extent that thye have in dicots. Little is known about the catalytic specificities of UGT92 and UGT89 UGTs. UGTs belonging to the UGT94 and UGT79 families are predominantly involved in glycosylation of sugar moieties. The absence of sequences representing these UGT families within the *A. strigosa* EST collection may reflect the absence of genes for these families from the genome. Alternatively such genes may be present within the genome but not expressed in the root tips.

The metabolite composition of oat tissues is not well documented. Nevertheless, in addition to avenacin, few other glycosides have been reported in oat root. The presence of flavone C-glycosides in root and leaves has been reported (Soriano et al. 2004). Scopoletin-7-*O*-glucoside (or scopolin) is exuded from oat root (Fay and Duke 1977). Blumenin a diglycoside of the terpenoid blumenol C accumulates in response to arbuscular mycorrhizal colonisation (Maier et al. 1997; Maier et al. 1995). The diversity of UGT families represented in the collection suggests that oat UGTs are involved in a variety of processes in oat root tips.

Table 3.2: Full-length *A. strigosa* UGTs and their closest enzyme functionally characterized.

Oat UGTs	Percentage of Identity	UGT with known activity	groups	Activity	Acc Number	Publication
AsGT00260	54%	UGT85A24	O	Iridoid glucosyltransferase	AB555732	Nagatoshi et al. 2011
AsGT08947	55%	UGT85A24	O	Iridoid glucosyltransferase	AB555732	Nagatoshi et al. 2011
AsGT06751	44%	UGT72B14	E	Tyrosol glucosyltransferase	EU567325	Yu et al. 2011
AsGT02436	39%	UGT73C6	D	Brassinosteroid 23-O-glucosyltransferase	AEC09298	Husar et al. 2011
AsGT01092	46%	UGT88A4	E	Coumarin-O-glucosyltransferase	ABL85471	Tian et al. 2006
AsGT06492	49%	UGT88A4	E	Coumarin-O-glucosyltransferase	ABL85471	Tian et al. 2006
AsGT23781	44%	UGT73C12	D	Hederagenin 3-O-glucosyltransferase	AFN26668	Augustin et al. 2012
AsGT23818	54%	UGT706D1	E	Flavonoid 3(7)-O-glucosyltransferase	BAB68093	Ko et al. 2008
AsGT23586A	51%	UGT706D1	E	Flavonoid 3(7)-O-glucosyltransferase	BAB68093	Ko et al. 2008
AsGT23586B	52%	UGT706D1	E	Flavonoid 3(7)-O-glucosyltransferase	BAB68093	Ko et al. 2008
AsGT01989	53%	UGT13248	L	Deoxynivalenol 3-O-glucosyltransferase	GU170355	Schweiger et al. 2010
AsGT18535	36%	UGT83A1	I	ND	Q9SGA8	Ross et al. 2001
AsGT16525	66%	UGT707A3	E	Flavonol 3-O-glucosyltransferase	BAC83989	Ko et al. 2008
AsGT05827	67%	cisZog1	O	Cis-zeatin-O-glucosyltransferase	AAK5355	Martin et al. 2001
AsGT18279	33%	GmSGT3	A	Soyasaponin-O-rhamnosyltransferase	BAI99585	Shibuya et al. 2010
AsGT17576	46%	ant5GT	L	Anthocyanin 5-O-glucosyltransferase	BAD06874	Imayama et al. 2004
AsGT01670	48%	Bx9	H	Benzoxazinoid-O-glucosyltransferase	AAL57038	von Rad et al. 2001
AsGT03158	44%	ant5GT	L	Anthocyanin 5-O-glucosyltransferase	BAD06874	Imayama et al. 2004
AsGT10326	53%	UGT86A1	K	ND	Q9SJL0	Ross et al. 2001
AsGT01332	36%	GmSGT3	A	Soyasaponin-O-rhamnosyltransferase	BAI99585	Shibuya et al. 2010
AsGT00733	41%	UGT90A7	C	flavonoid 7,4-O-GlcT	ACB56926	Witte et al. 2009
AsGT26962	44%	UGT87A1	J	ND	O64732	Ross et al. 2001
AsGT1a15	50%	FaGT2	L	Cinnamate-O-glucosyltransferase	Q66PF4	Landmann et al. 2007
AsGT3i21	36%	RuGT-5	D	Flavonoid-O-glucosyltransferase	XM_463383	Ko et al. 2006
AsGT4h2	54%	Bx9	H	Benzoxazinoid-O-glucosyltransferase	AAL57038	von Rad et al. 2001
AsGT5d1	63%	cisZog1	O	Cis-zeatin-O-glucosyltransferase	AAK5355	Martin et al. 2001

Table 3.2: Full-length *A. strigosa* UGTs and their closest enzyme functionally

Oat UGTs	Percentage of Identity	UGT with known activity	groups	Activity	Acc Number	Publication
AsGT84	55%	UGT85A24	G	Iridoid glucosyltransferase	AB555732	Nagatoshi et al. 2011
AsGT11i1	34%	RuGT-5	D	Flavonoid-O-glucosyltransferase	XM_463383	Martin et al. 2001
AsGT12o13	65%	UGT709A4	G	Isoflavonoid 7-O-glucosyltransferase	BAC80066	Ko et al. 2008
AsGT14h20	40%	UGT73C6	D	Brassinosteroid 23-O-glucosyltransferase	AEC09298	Husar et al. 2011
AsGT14h21	40%	UGT73C12	D	Hederagenin 3-O-glucosyltransferase	AFN26668	Augustin et al. 2012
AsGT15a11	47%	Gtsatom	L	Gentisate(SA)-O-xylosyltransferase	CAI62049	Tarraga et al. 2005
AsGT16f23	45%	UGT73C10	D	Hederagenin 3-O-glucosyltransferase	AFN26666	Augustin et al. 2012
AsGT16h6	71%	UGT85B1	G	Cyanohydrin-O-glucosyltransferase	Q9SBL1	Hansen et al. 2003
SAD10	47%	Gtsatom	L	Gentisate(SA)-O-glucosyltransfearse	CAI62049	Tarraga et al. 2005
AsGT20e13	40%	ant5GT	L	Anthocyanin 5-O-glucosyltransferase	BAD06874	Imayama et al. 2004
AsGT20n10	47%	UGT88aA4	E	Coumarin-O-glucosyltransferase	ABL85471	Tian et al. 2006
AsGT21p16	55%	UGT85A24	G	Iridoid glucosyltransferase	AB555732	Nagatoshi et al. 2011
AsGT24i2	72%	RUGT-5	D	Flavonoid-O-glucosyltransferase	XM_463383	Ko et al. 2006
AsGT24a3	68%	cisZog1	O	Cis-zeatin-O-glucosyltransferase	AAK5355	Martin et al. 2001
AsGT25n16	56%	UGT706C1	E	Flavonoid 3-O-glucosyltransferase	BAB68090	Ko et al. 2008
AsGT27a12	65%	cisZog1	O	Cis-zeatin-O-glucosyltransferase	AAK5355	Martin et al. 2001
AsGT27f7	41%	UGT73C6	D	Brassinosteroid 23-O-glucosyltransferase	AEC09298	Husar et al. 2011
AsGT28b19	45%	GT45	L	Flavonoid-O-glucosyltransferase	Q6X1C0	Moraga et al. 2009
AsGT29m5	47%	GTsatom	L	Gentisate(SA)-O-xylosyltransferase	CAI62049	Tarraga et al. 2005
AsGT30a8	47%	Bx8	H	Benzoxazinoid-O-glucosyltransferase	AAL57037	von Rad et al. 2001
AsGT24138	93%	AsSGT1		Sterol glycosyltransferase	CAB06081	Warnecke and Heinz 1994
AsGT24525	90%	AsSGT1		Sterol glycosyltransferase	CAB06081	Warnecke and Heinz 1994
AsGT03999	63%	UGT80B1		Sterol glycosyltransferase	AEE31976	DeBolt et al. 2009
AsGT00678	75%	UGT81A1		monogalactosyldiacylglycerol synthase	AL031004	Jorasch et al. 2000

3.2.3   Features of oat UGTs

3.2.3.1   *Essential catalytic residues are conserved in A. strigosa UGTs*

Several amino acid residues have been shown to be important for catalysis of sugar transfer following a SN2-like mechanism. A conserved histidine (His) at the N-terminal end of UGTs plays the role of a general base. An aspartate (Asp) residue is believed to form hydrogen bonds with this His residue to stabilize it (Lairson et al. 2008; Osmani et al. 2009; Wang 2009). The catalytic His residue is conserved in all *A. strigosa* UGT sequences identified in this work (Fig. 3.2.B). Most of the *A. strigosa* UGTs possess the Asp residue; although it is absent from six of the UGT sequences. Five of these UGTs belong to either the UGT75 or UGT84 families (Fig. 3.2.B). Functionally characterised enzymes from these families also lack this Asp. Therefore an alternative mechanism may exist in those families. Intriguingly, AsGT02436 possesses a glycine (Gly) instead of the Asp residue, in contrast to other active UGT73s (Fig. 3.2.B). Conservation of these catalytic residues within the *A. strigosa* UGT collection suggests that most of our collection is composed of potentially active enzymes.

3.2.3.2   *Analysis of the PSPG motif suggests that the majority of the oat UGTs are glucosyltransferases*

Plant UGT phylogeny reconstruction can provide clues about potential acceptors of oat UGTs but offers very little information about sugar donor preferences. Phylogenetic analysis of active UGTs highlights the fact that the ability to specifically recognise a UDP-sugar has evolved independently multiple times (Fig. 3.1). Members of a single UGT subfamily can perform transfer of several different monosaccharides (see section 1.2.5). However, most plant UGTs have a strong preference for a single UDP-sugar, as revealed by comparative kinetic analysis performed on multiple sugar donors (Kubo et al. 2004; Noguchi et al. 2009; Sayama et al. 2012). Residues responsible for sugar donor recognition have been the centre of much interest in the past decade and the mechanisms underlying sugar donor specificity are beginning to be understood.

The PSPG motif shapes the sugar-donor binding site of UGTs and consequently is primarily involved in UDP-sugar specificity (Wang 2009). The PSPG motif was identified in all of *A. strigosa* UGT sequences after alignment with characterised UGT71G1 (Fig. 3.2.B). To investigate amino residue conservation within the PSPG motif of *A. strigosa* UGTs compared to those from the dicot species *A. thaliana*, a graphical representation of the consensus sequences were generated from the available UGT sequences for each species (Fig. 3.2.A). Crystal structures of plant UGTs and directed mutational approaches have helped to identify crucial residues involved in interaction with the UDP moiety of the sugar donor (Osmani et al. 2009). These are marked with an asterisk on both consensus motifs. These residues are generally conserved between *A. thaliana* and *A. strigosa* (Fig. 3.2.A-B). The overall similarity between the two consensus motifs suggests that interactions between UGTs and sugar nucleotides are conserved between monocots and dicots, and that the sugar donor binding sites appear likely to be functional in the majority of the *A. strigosa* UGTs identified.

Intriguingly, some of the residues known to interact with the nucleotide diphosphate moiety of the donor are not conserved in the PSPG sequences of AsGT11i11 (His19), AsGT27f7 (Ser24), AsGT02436 (Asn23) and AsGT01332 (Asn23) (Fig. 3.2.B). None of the enzymes included in the phylogenetic tree (Fig. 3.1) has His19 substituted by another residue. Interestingly, substitution of Asn23 with glycine appears to be associated with rhamnosyltransferase activity in several cases: UGT89C1, CmRhaT and GmSGT3 (Frydman et al. 2013; Shibuya et al. 2010; Yonekura-Sakakibara et al. 2007); additionally Asn23 is generally not conserved in GGTs. The rare substitutions occurring for critical residues involved in uridine diphosphate binding may affect sugar donor recognition of the corresponding enzymes.
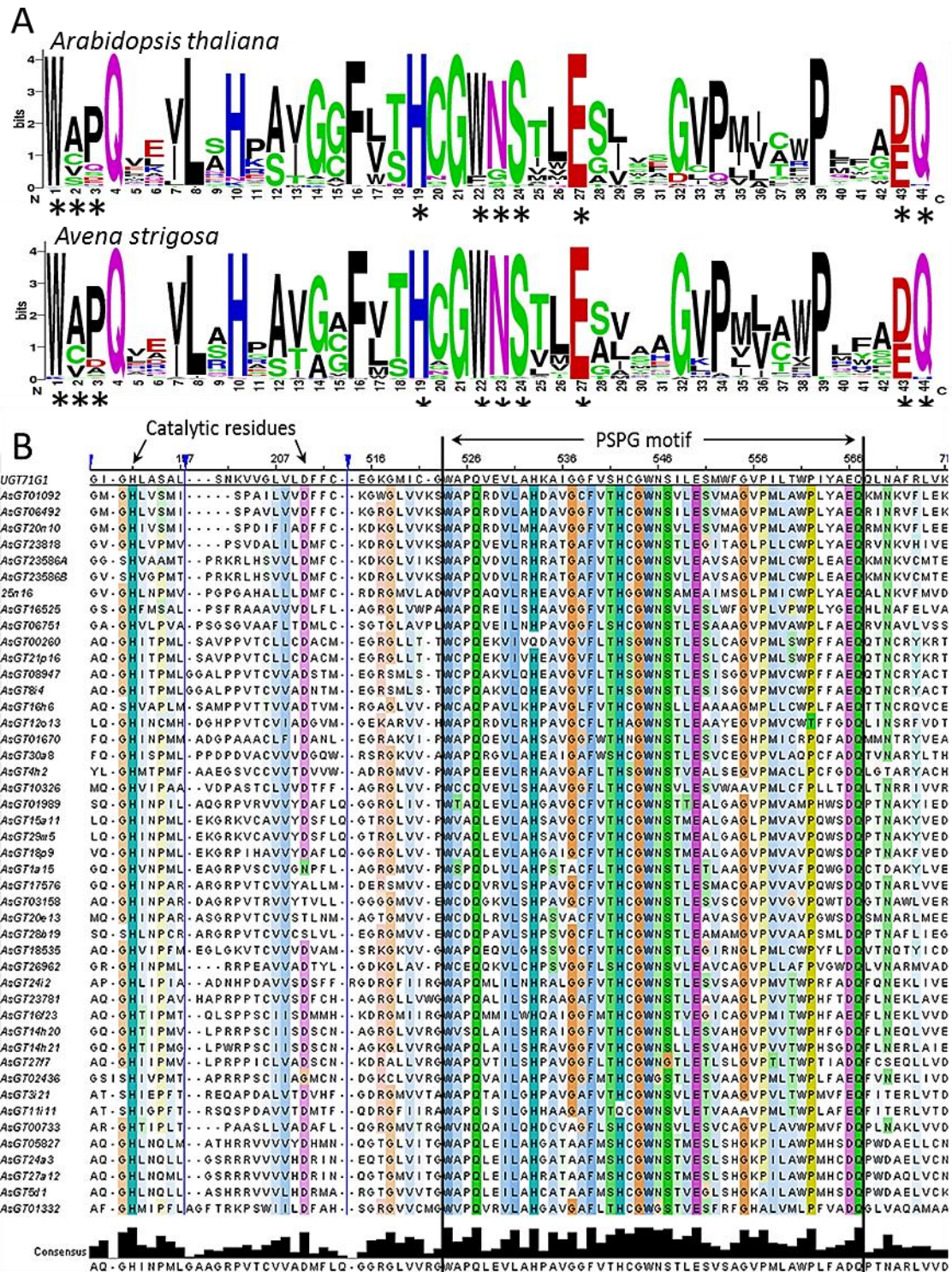
Figure 3.2: Amino acid sequence conservation of *A. strigosa* UGTs. A: Comparison of the PSPG consensus motifs derived from UGTs from *Arabidopsis thaliana* (top) and *Avena strigosa* (bottom). *A. thaliana* sequences were retrieved from http://www.p450.kvl.dk/UGT.shtml. A graphical representation of residues constituting the PSPG motifs was made using weblogo.berkeley.edu software. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. Essential residues for UDP-sugar binding are labelled with an asterisk. **B:** Alignment of *A. strigosa* full-lenght UGTs with the sequence of the crystallised UGT71G1 (Shao et al. 2005). The alignment shows highly conserved regions of UGTs; the PSPG motif and essential catalytic residues are highlighted in red. The consensus sequence of the alignment shows conserved residues identified in crystallographic studies. Key residues for activity or sugar specificity are labelled with an asterisk.

Some residues of the PSPG motif that are known to interact directly with the glycosidic moiety of the sugar donor UDP-Glc (Trp22, Asp/Glu43 and Gln44; refs) are conserved in *A. strigosa* UGTs (Fig. 3.2.A-B). The absence of variation in these residues (especially PSPG terminal position 43rd and 44th; see section 1.2.3) may indicate that the majority of the enzymes in our collection are glucosyltransferases. Residues This assumption has to be tempered by the fact that other regions in the *C*- and *N*-terminal domains of UGTs have also been suggested to interact with sugar donors (Osmani et al. 2009). The fact that targeted mutagenesis of some of the residues within the PSPG motif that are implicated in sugar binding has often failed to produce a change of sugar specificity also suggests that discrimination between UDP-sugars is not dependent solely on the PSPG motif (Kubo et al. 2004; Noguchi et al. 2009).

3.2.4 The role of Group D enzymes in triterpenoid glycosylation and the expansion of group D in monocots

As mentioned before, group D (which consists of members of family UGT73) is particularly rich in glycosyltransferases that are active towards triterpenes and related compounds in other plants (Fig. 3.3, blue accessions). A phylogenetic tree was then built that included all functionally characterised UGT73s plus the entire UGT73 family of the sequenced monocot species *O. sativa* and of the dicot species *A. thaliana.* This phylogenetic tree enabled comparison of the relatedness between monocot and dicot UGT73 sequences (Fig. 3.3).

3.2.4.1 *Relation between functions and phylogeny of group D glycosyltransferases*

As observed for other UGT families (Caputi et al. 2011), the members of the UGT73 family are divided into monocot- and dicot-specific branches (Fig. 3.3, black and grey branches). These have been designated D1 and D2 for the dicot sequences and M1, M2, M3, M4 and M5 for the monocot sequence (Fig. 3.3). All UGT73 enzymes that have been functionally characterised so far are from dicot species except the flavonoid glucosyltransferase RUGT-5 (Ko et al. 2006). The dicot UGT73s are separated into two major clusters D1 (sub-families UGT73C, D, E and N) and D2 (sub-families UGT73A, B, F, K and P). There is no obvious functional segregation between those clusters. Triterpenoids, steroids and flavonoids are used as acceptors by enzymes from both the D1 and D2 groups. The group D is known to

form the cluster IIIa of flavonoid glycosyltransfersases and enzymes from this group are generally presented as catalysing regiospecific addition of glycoside onto C-7 position of flavonoid acceptors (Frydman et al. 2013; Noguchi et al. 2009; Yonekura-Sakakibara and Hanada 2011). Triterpenoid glycosyltransferases repartition inside group D (Fig. 3.3, blue accessions) and over the whole plant UGT phylogeny suggests such phylogenetic clustering with conserved regiospecific glycosylation did not occur for triterpenoid glycosyltransferases. Sayama et al. (2012) suggested that the structural similarity displayed by a majority of triterpenoid glycosyltransferases (the one from group D) regardless of their regiospecificity might indicate a recent evolution of triterpenoid UGTs from an ancestral UGT and then evolution of current glycosylation positions.

3.2.4.2 *Oat glycosyltransferases from group D, potential triterpenoid glycosyltransferases*

Interestingly, AsGT16f23 is part of group M1 orthologous to group D1 containing triterpenoid-3-*O*-glycosyltransferases from *B. vulgaris* and brasinosteroid-23-*O*-glycosyltransferases from *A. thaliana* (Augustin et al. 2012; Husar et al. 2011; Poppenberger et al. 2005). Common functions may therefore be conserved between M1 and D1. AsGT24i2 displays 72% identity with the flavonoid-3-*O*-glucosyltransferase RUGT-5 from rice in group M5 (Ko et al. 2006). The absence of functionally characterised monocot UGT73 enzymes prevents speculation about the potential activities of the remaining *A. strigosa* UGT73 enzymes. Functional characterisation of these enzymes will shed light on the role of family UGT73 in monocots and in higher plants in general. This phylogenetic analysis and recent reports on genome-wide analysis of plant UGT families (Caputi et al. 2011; Yonekura-Sakakibara and Hanada 2011) suggests a larger expansion and diversification of UGT73 enzymes in monocot species compare to dicots. This might suggest the development of distinct functions in monocot UGT73 family.
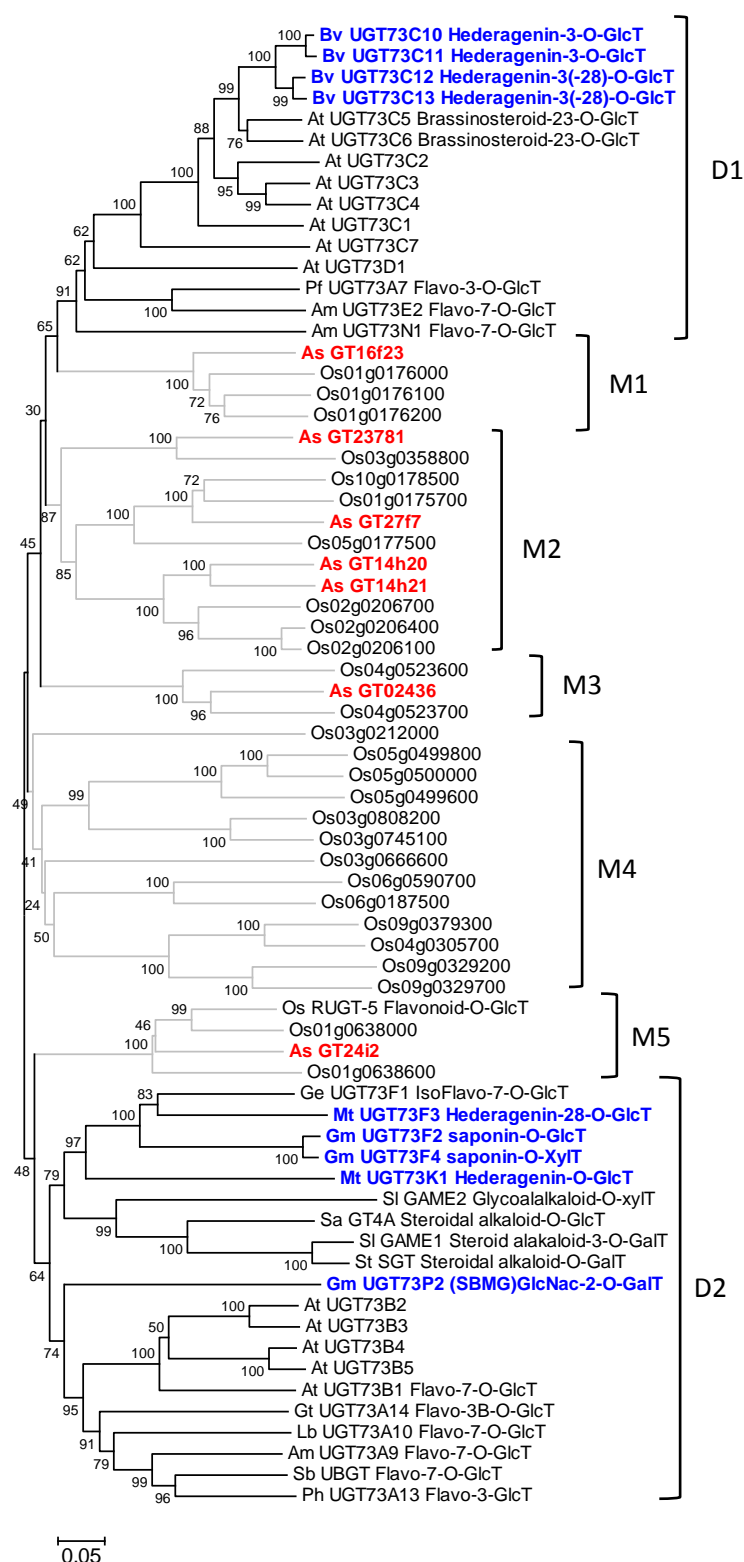
Figure 3.3: Phylogenetic tree of plant UDP-glycosyltransferases belonging to group D.
The deduced amino acid sequences of the *A. strigosa* UGTs (shown in red) were aligned with biochemically characterised UGT73 enzyme. All the UGT73 sequences from *O. sativa* and *A. thaliana* were included in the tree. Monocot specific branches are drawn in light grey. Tritepenoid UGTs are indicated in blue. The phylogenetic tree was drawn using the Neighbor-Joining method with 500 bootstrap replicates. Further information about the sequences in this phylogenetic tree are provided in supplement (Supp. S.3 and S.5). The scale bar represents 5% divergence.

3.2.4.3 *Triterpenoid glycosyltransferase candidates out of group D*

In addition to the UGT73 family, sterol glycosyltransferases may also be potential candidates for avenacin glycosylation considering the recruitment of SAD1 and SAD2 enzymes from the primary sterol pathway to avenacin biosynthesis. The two glucosylation steps required for avenacin trisaccharide formation (β1-2 and β1-4 glucosylation of arabinose) requires GGT activities. As mentioned before GGTs are generally belonging to a large branch of UGT phylogeny comprising families UGT79, UGT91 and UGT94; AsGT01332 and AsGT18279 belong to this branch and are potential candidates for avenacin glucosylation processes (Fig. 3.1).

3.2.5 Analysis of the *A. strigosa* root proteome

3.2.5.1 *Proteomic analysis reveals differential representation of UGTs in the root tips and elongation zone*

Previously, immunoblot analysis has shown that the SAD proteins accumulate specifically in the root tips of *A. strigosa* seedlings (Mugford et al. 2013; Owatworakit et al. 2012). As a complementary strategy to investigate the UGTs present in the roots of *A. strigosa* seedlings, proteomic analysis was carried out using linear trap quadrupole (LTQ)-orbitrap. The LTQ-orbitrap offers exact masses of the orbitrap mass analyser - crucial for the rigorous identification of tryptic peptides - associated with efficiency and sensitivity of the linear ion trap. Recently orbitrap mass spectrometry has been used with complex protein mixtures extracted from plant tissues to great effect. In rice, large-scale proteomic analysis was used to generate a database representing the products of 3200 genes, so allowing refinement of rice genome annotation (Helmy et al. 2011). Proteomic analysis of Holm oak revealed qualitative and quantitative variation of the protein profiles of pollen from various origins (Valero Galvan et al. 2012).

Proteins were extracted from the root tips and elongation zones of five-day old *A. strigosa* seedlings (Fig. 3.4.A). Protein quantification was carried out using Bradford reagent and the protein solutions were diluted to 0.32 mg/ml each. LTQ-Orbitrap analysis was carried out by G. Saalbach (John Innes Centre Proteomics platform). This included tryptic digestion of the protein samples and quantification/identification of detected proteins using MaxQuant and Mascot software. The search was performed on the TAIR protein database (*A. thaliana*)

augmented with the entire collection of UGT sequences obtained from *A. strigosa* transcriptomic analysis.

Analysis of the total protein extract from the root tips resulted in identification of AsGT11i11 and AsGT16h6 only. This poor coverage was likely to be due to saturation of the Orbitrap detector, resulting in the detection of only a few peptides. To simplify the sample mixture we therefore took advantage of the conserved length of UGTs; *A. strigosa* UGT molecular weights are between 48kDa and 55kDa. Proteins with molecular weights 45-57 kDa were extracted from gel slices and the mixture injected into the Orbitrap following tryptic digestion (Fig. 3.4.B). Analysis of root protein extracts restricted to the targeted molecular weight window was more informative. SAD10 was unambiguously identified from root tip and root elongation zone based on Orbitrap analysis of four peptides (Fig. 3.4.C).

Differences in the abundance of the SAD10 protein were observed when extracts from the root tip and elongation zone were compared, with higher SAD10 protein levels in the root tips (Fig. 3.4.C). This is consistent with the role of SAD10 in avenacin biosynthesis (Owatworakit et al. 2012). A similar pattern of protein accumulation was observed for the SAD10 homologue, AsGT15a11 (UGT74H6). This is consistent with the likely role of UGT74H6 in biosynthesis of the non-fluorescent avenacins, A-2 and B-2 (Owatworakit et al. 2012). Several other UGTs have a similar distribution in oat root tissues, for example AsGT11i11, AsGT23586, AsGT23781, AsGT16525 and AsGT10326. AsGT1a15, AsGT20n10, AsGT05827 and AsGT10326 were detected almost exclusively in the root tips (Fig. 3.4.C). The differential distribution of these UGTs in *A. strigosa* roots suggests that they are involved in processes that are specific to root tip. Interestingly, the only UGT73 enzyme that showed a similar distribution to SAD10 is AsGT23781, part of the M2 cluster from which function remains unknown (Fig. 3.4). AsGT16h6, AsGT17576 and AsGT01989 have a different distribution in roots, with higher levels of protein detected in elongation zone. Similar levels of protein were detected in both parts of the root for AsGT21p16 (Fig 3.4.C).

This proteomic analysis was done with one sample from each part of the root and so the analysis would need to be repeated in order to ensure that these findings are reproducible. Nevertheless this preliminary analysis provides a valuable complement to the transcriptomic analysis.
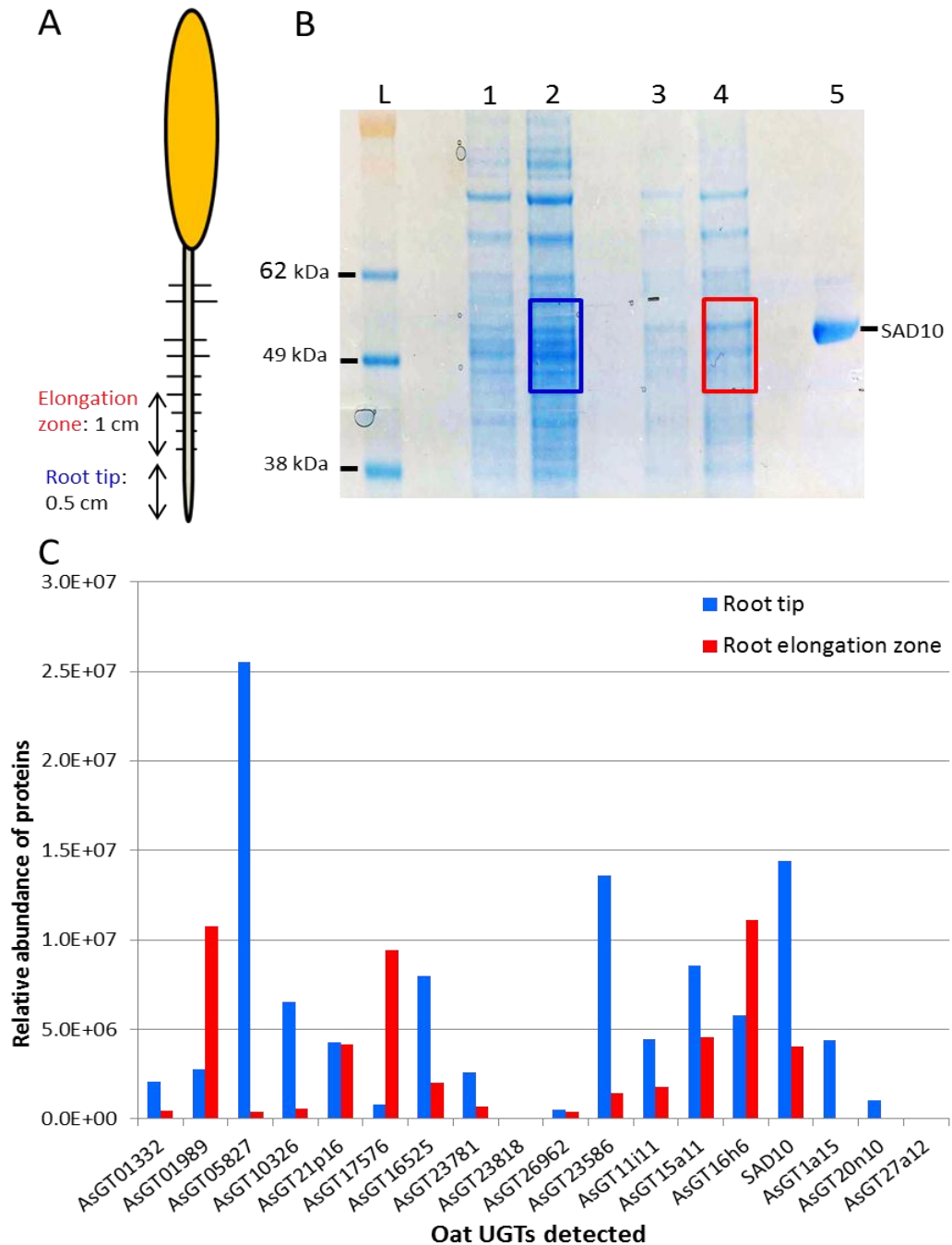
Figure 3.4: Proteomic analysis of *A. strigosa* UGTs. A. Proteins were extracted from 3-day-old seedlings of *A. strigosa*. A total of 50 mg of tissue was collected from the elongation zone (EZ) and root tips (RT). **B.** SDS-PAGE gel slices corresponding to the molecular weight of UGTs were used in LTQ-Orbitrap analysis. Lane L, SeeBlue® Plus2 Pre-Stained Standard; lane 1, RT protein extract 0.32 mg/mL; lane 2, RT protein extract 0.15 mg/mL, lane 3, EZ protein extract 0.32 mg/mL; Lane 4, EZ protein extract 0.15 mg/mL; Lane 5, recombinant SAD10. **C.** Comparison of relative abundance of *A. strigosa* UGTs between RT and EZ tissues. Orbitrap mass spectrometry was performed on trypsic products of proteins from gel slices shown above. Relative intensities of peptides from *A. strigosa* RT are in blue; relative intensity of peptides from the EZ are in red. SAD10 is part of the avenacin biosynthetic pathway.

3.2.5.2 *Comparison of UGTs expressed in root tip of wild-type, sad3 and sad4 mutants*

The *sad3* and *sad4* are affected in glycosylation of avenacin and accumulate mono deglucosylated avenacin. The *sad3* and *sad4* mutants display a severe phenotype affecting root development and vesicular trafficking in root cells (Mylona et al. 2008). In order to analyse the *sad3* and *sad4* mutant proteome without introducing a bias due to severe cellular defects we used oat lines with a *sad1* mutant background. The *sad1* mutant is affected in the first step of avenacin biosynthesis; it is unable to make cyclisation of 2,3-oxidosqualene leading to an early shut-down of the whole avenacin pathway. In a *sad1* background *sad3* and *sad4* mutants have a wild-type phenotype (Mylona et al. 2008). Total protein extract from wild-type oat together with *sad1* mutant and double mutant lines *sad1*/*sad3* and *sad1*/*sad4* mutant were processed for Orbitrap analysis of gel slices. No major differences in UGTs expression profile were observed in any of the mutant lines. No mutation was detected in UGT peptides sequences identified from this analysis.

The resolution obtained from the mass-spectrometry analysis did not allow full coverage of UGTs of interest. Therefore only a complete loss of the protein or a mutation in one of the detected peptides may be identified using the Orbitarp analysis. Proteomic analysis of *sad3* and *sad4* mutant may suggest their phenotype is not directly linked to UGT but we cannot exclude a role of one of the detected UGT in phenotype of *sad3* or *sad4* mutants.

A large scale analysis of all UGTs present in the oat collection was possible due to the Orbitrap proteomic analysis. SAD10 was unambiguously detected and its histological pattern was consistent with its role in avenacin biosynthesis. Detection of SAD10 and UGT74H7 suggests resolution of the Orbitrap is appropriate to identify proteins involved in the avenacin pathway. Several UGTs of the collection co-localised with SAD10 and might be considered as potential avenacin glycosyltransferase candidates.

3.2.6   <u>Gene expression analysis reveals that some *A. strigosa* UGTs are co-expressed with Sad genes</u>

Biosynthetic genes of secondary metabolites are generally expressed in a particular tissue (Schilmiller et al. 2012). Consequently, gene expression analysis is a powerful way to uncover enzymes involved in the biosynthetic pathways of a targeted secondary metabolite. Investigation of coordinated regulation of genes in tissues accumulating a given metabolite have been used successfully to uncover biosynthetic genes of a variety of compounds like flavonoids, indoles, phenylpropanoids (Gachon et al. 2005a; Pang et al. 2008; Yonekura-Sakakibara et al. 2012; Yonekura-Sakakibara et al. 2008). Co-expression analysis of tailoring genes with OSCs was a successful approach to discover UGTs and P450s involved in saponin biosynthesis in *Medicago truncatula* (Achnine et al. 2005; Naoumkina et al. 2010). The *Sad1*, *Sad2*, *Sad7* and *Sad9* genes are all expressed preferentially in the root tips with little or no expression in other tissues (Fig 1.8) (Haralampidis et al. 2001; Mugford et al. 2013; Qi et al. 2006). Genes encoding enzymes involved in glycosylation of the triterpene scaffold during avenacin biosynthesis may therefore be expected to have a similar expression pattern.

Expression of *Ugt* genes were analysed by mRNA-reverse transcription-PCR (RT-PCR). cDNA was generated from the following parts of 3-day-old *A. strigosa* seedlings: root (R), root tip (RT), root elongation zone (EZ) and young leaf (YL) (Fig. 3.5.A), and from the following parts of older plants: mature leaf (ML), flower (F) and stalk (S) (Fig. 3.5.A). The housekeeping gene encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and the previously characterised avenacin biosynthetic gene *Sad10* were used as controls. Amplification of the *Sad10* transcript revealed a root tip-specific expression profile in accordance with previous publications (Fig. 3.5.B). Specific primers were designed for UGT genes presenting characteristics of avenacin glycosyltransferases candidates based mainly on phylogeny and protein levels in oat root (see table 3.3). Specifically, UGTs from phylogenetic groups of interest (group D and sterols UGTs) and UGTs showing high protein accumulation levels in root tips or higher accumulation in root tips than elongation zone were picked up for the present analysis. Optimised PCR conditions were established to amplify the desired fragment from each transcript.
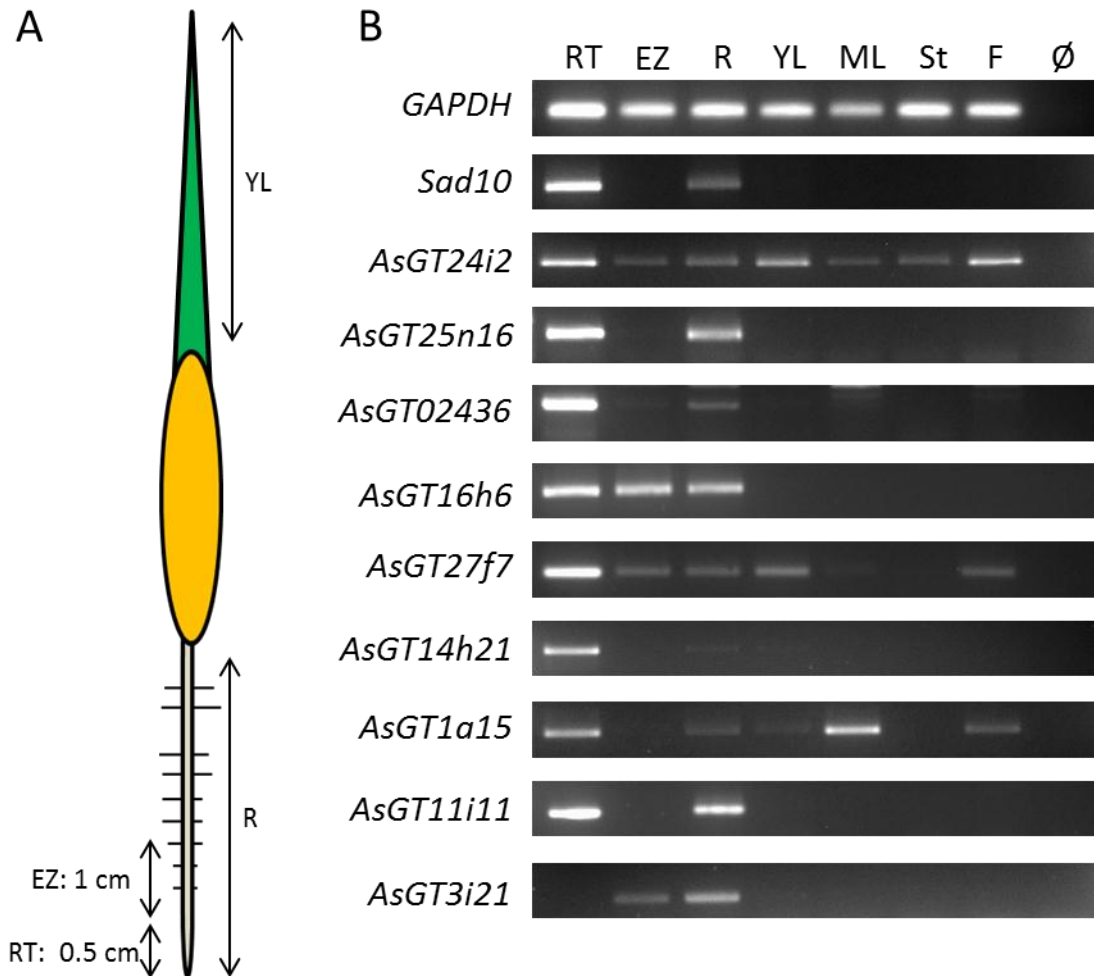
Figure 3.5: Expression analysis of UGT genes in *A. strigosa* tissues. A. Expression analysis was conducted using mRNA-reverse transcription-PCR (RT-PCR) technique. *A. strigosa* tissues used were from 3-days-old seedlings (RT: root tip, RE: root elongation zone; R: entire young root, YL: young leaf) or tissues of flowering plants (ML: mature leaf, St: stalk and F: flower). **B.** Examples of expression profiles obtained from RT-PCR DNA over the seven tissues plus negative control (Ø). The negative control consists in PCR reaction mixture without cDNA. The *Sad10* gene is part of the avenacin cluster from which all genes (*Sad1, Sad2, Sad7, Sad9* and *Sad10)* have been shown to be tightly co-expressed. Positive control consists on amplification of glyceraldehyde 3-phosphate dehydrogenase (GAPDH).

Similarly to other *Sad* genes *Sad10* is tightly co-expressed with *Sad1* (Supp. S.6). Two other genes that are closely related to *Sad10* - *AsGT29m5* and *AsGT15a11* (*Ugt74H6* and *UGT74H7*) - are also expressed exclusively in the root tips. This gene expression analysis supported the idea that AsGT29m5 and AsGT15a11 may participate in the synthesis of avenacins by glucosylating acyl donors required as substrates for SAD7 activity (Fig. 3.6) (Owatworakit et al. 2012); UGT01989 is part of the monophyletic branch containing SAD10, UGT29m5 and UGT15a11; its activity is unknown. In contrast to the other three *Ugt* genes *AsGT01989* is expressed in the elongation zone but not in the root tips (table 3.3).
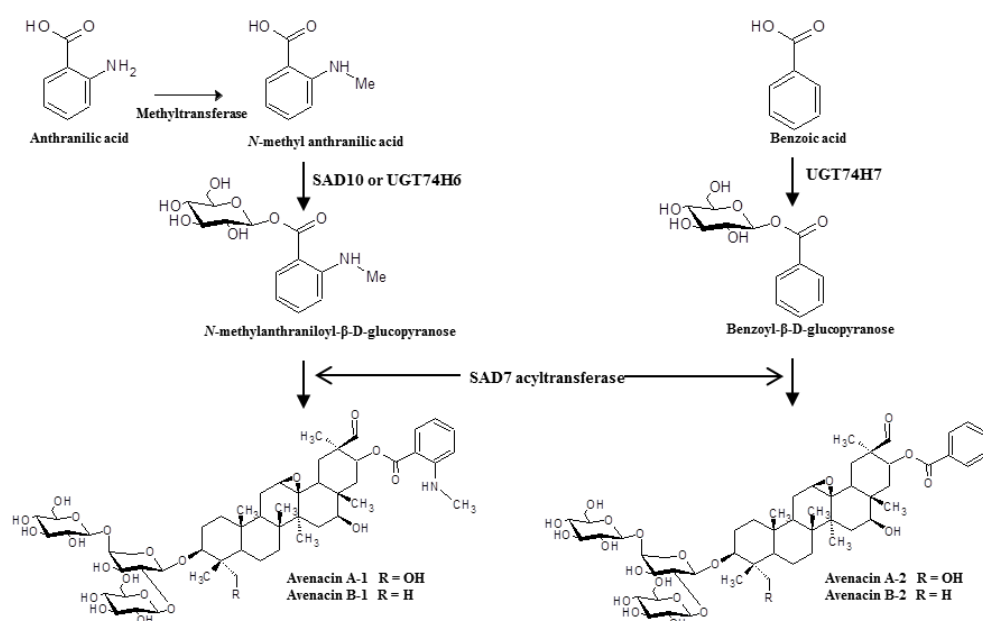
Figure 3.6. Predicted function of SAD10, AsGT29m5 (UGT74H6) and AsGT15a11 (UGT74H7) in avenacin biosynthesis (Owatworakit et al. 2012).

As mentioned before UGTs from group D (or UGT73s) are of particular interest with regard to identifying triterpene glycosyltransferases (see section 3.2.4). The expression profiles of all UGT73s within the collection were determined by RT-PCR. Interestingly, *AsGT02436* and homologous genes *AsGT14h20* and *AsGT14h21* display a *Sad* gene-like expression profile, their transcripts being detected exclusively in the root tips. Like SAD10, the AsGT23781 protein accumulates principally in the root tips. In agreement with the proteomic analysis, the corresponding transcript was detected in the root tips and also in flowers. *AsGT24i2*, *AsGT16f23* and *AsGT27f7* are expressed in various tissues (table 3.3).

Proteomic analysis identified several *A. strigosa* UGTs with a similar differential display than SAD10 in oat roots (Fig. 3.4; section 3.2.5.1). Such a feature is of interest for the finding of avenacin glycosyltransferases candidates. Beyond those genes only *AsGT11i11* is strictly expressed in root tips. As mentioned previously *AsGT23781* is expressed in root tip and flowers. Homologous genes *AsGT23786A* and *AsGT23786B* are predominantly expressed in root tip but low expression levels are also detected in elongation zone and leaf tissues. Expression profiles of the remaining genes are not correlated with the characterised *Sad* genes (table 3.3). The proteomic analysis also suggests a substantial accumulation of AsGT21p16 and AsGT16h6 proteins in *A. strigosa* roots. Interestingly, *AsGT21p16* gene is essentially expressed in root tip and *AsGT16h6* is expressed only in roots (table 3.3).

Sterol UGT80s may be of interest for avenacin biosynthesis considering the evolution of *Sad1* and *Sad2* from sterol metabolism. The three sterol UGTs identified in our collection (*AsGT24525*, *AsGT03999* and *AsGT24138*) are all expressed in various tissues and therefore, are unlikely to contribute to avenacin synthesis (table 3.3). *AsGT3i21* is closely related to *AsGT11i11*, both enzymes are distant from any other characterised UGTs (Fig. 3.1), functional investigation on those enzymes is therefore of particular interest. The expression of these two genes seems mutually exclusive as *AsGT3i21* is expressed in elongation zone but not in root tip and *AsGT11i11* expression is restricted to root tip. This may suggest both enzymes are involved in glycosylation of different acceptors or used different sugar donors to decorate the same acceptor. *AsGT25n16* was reported to be expressed exclusively in root tip from preliminary experiment on oat UGTs (Osbourn lab unpublished data), the present gene expression profile analysis confirmed it.

A majority of the UGT genes analysed above are expressed in oat root tips; this is proving the relevance of the initial transcriptomic approach for identification of tailoring enzymes potentially involved in root secondary metabolism. Several UGTs included in the analysis are co-expressed with the *Sad* genes and may be prioritised candidates for avenacin glycosylation. UGT genes not strictly expressed in root tip also need to be considered carefully; really some UGTs have been reported to have several functions in a single plant species. For example, two independent studies have shown evidence of a role of UGT73C6 in brassinosteroid

and flavonoid metabolism in *A. thaliana* (Husar et al. 2011; Jones et al. 2003). If avenacin glycosyltransferases are involved in other mechanisms they may be expressed in other tissues as well as in the root tips. The expression profiles generated during this study are summarized in (table 3.3).

A. strigosa UGT genes	Reason for selection as candidate gene	Expression profiles						
		RT	EZ	YR	YL	ML	St	F
Sad10	control	+++	-	++	-	-	-	-
AsGT15a11	*Sad10* homologue	+++	-	++	-	-	-	-
AsGT29m5	*Sad10* homologue	+++	-	++	-	-	-	-
AsGT01989	*Sad10* homologue	-	+	+++	-	-	-	-
AsGT14h20	Group D	++	-	+	-	-	-	-
AsGT14h21	Group D	+++	-	+	+	-	-	-
AsGT02436	Group D	+++	+	++	-	-	-	-
AsGT16f23	Group D	+++	+	++	++	+	+	+
AsGT27f7	Group D	+++	++	++	++	+	-	+
AsGT24i2	Group D	+++	++	++	+++	++	++	+++
AsGT23781	Group D	++	-	++	-	-	-	++
AsGT23586A	protein level in RT	+++	++	+++	-	+	-	-
AsGT23586B	protein level in RT	+++	+	+++	+	+	-	+
AsGT05827	protein level in RT	+++	+	+++	+++	+	-	++
AsGT11i11	protein level in RT	+++	-	++	-	-	-	-
AsGT16525	protein level in RT	+++	++	+++	++	++	-	++
AsGT01a15	protein level in root	++	-	++	+	+++	-	++
AsGT21p16	protein level in root	+++	+	++	-	-	-	-
AsGT16h6	protein level in root	+++	+++	+++	-	-	-	-
AsGT25n16		+++	-	++	-	-	-	-
AsGT03i21	unclassified	-	++	++	-	-	-	-
AsGT03999	UGT80	+++	+	++	++	++	+	+
AsGT24525	UGT80	+++	+	++	+	+	-	+
AsGT24138	UGT80	++	+	++	++	++	-	++

Table 3.3: Summary table of *A. strigosa* UGTs expression profiles. Expression profile was obtained by semi-quantitative RT-PCR on various oat tissues: RT: root tip; EZ: elongation zone of the root, YR: entire root, YL: young leaf, ML: mature leaf, St: stalk and F: flower. Comparative expression scale: +++: maximal expression, ++: moderate expression, +: low expression, -: no expression detected.

3.3   Conclusion

3.3.1   Glycosyltransferases from oat root tips – extent and features

A large collection of UGTs was retrieved from transcriptomic analysis of oat root tips. A total of 53 full-length UGT sequences were part of the 110 UGT-like sequences identified by tBLASTn searches. Subsequent analysis of *A. strigosa* UGTs focused on full-length UGTs. An overview of the selection strategy is presented in figure 3.7.

Analysis of deduced amino acid sequences of oat UGTs has revealed the conservation of essential residues of the active sites and sugar donor binding sites between dicot and monocot species. These data suggest that most of the sequences of the collection may encode active enzymes, and that the majority of them are likely to have glucosyltransferase activity based on the amino acid composition of their PSPG motif. The final residue of the PSPG motif has been shown to be involved in discrimination between glucose and galactose (Kubo et al. 2004). Plant UGT galactosyltransferases possess a conserved histidine residue at that position. This residue is conserved in the only plant UGT arabinosyltransferase so far functionally characterised in plants (Yonekura-Sakakibara et al. 2008). The authors suggest that the histidine residue has a similar role in recognition of UDP-Gal and UDP-Ara, considering the similar configurations of hydroxyl groups at C-2, C-3, and C-4 positions in β-L-arabinose and α-D-galactose. An arabinose residue is directly attached at the C-3 position of avenacin triterpene backbone (Fig. 1.7.A); arabinosyltransferase activity is therefore of great interest within the scope of identifying the enzymes involved in avenacin trisaccharide formation. None of the full-length UGT sequences of the *A. strigosa* collection encode for this particular histidine. This may suggest that this structural basis for recognition of UDP-Ara is not conserved throughout the whole plant UGT family. Mutational analysis focussing on sugar specificity alteration of plant UGTs combined with structural modelling suggests that recognition of a specific sugar donor is not attributable to a single residue (Kubo et al. 2004; Noguchi et al. 2009; Shao et al. 2005). In addition to the PSPG motif, Osmani et al. (2009) pointed out the role of the first loop in the C-terminal domain (C1-loop), the interdomain linker and the N-terminal domain in sugar donor recognition.

3.3.2 Phylogenetic analysis brings new insights into possible functional evolution
of monocot species

The phylogenetic analysis reveals a heterogeneous distribution of *A. strigosa* UGTs over the phylogenetic groups defined previously (Ross et al. 2001). The absence of several UGT families within the oat sequence collection may signify a complete absence of those genes within the oat genome or a non-expression of those genes in oat root tip. The absence of UGT78 (group F) is in accordance with genome-wide phylogenetic analysis conducted by Caputi et al. (2011) over 12 plant genomes comprising two monocot species (rice and sorghum). The vast majority of UGT78s are involved in flavonoid-3-*O*-glycosylation and constitute clade I of the flavonoid glycosyltransferases (Ford et al. 1998; Jones et al. 2003; Kim et al. 2012; Modolo et al. 2009; Sui et al. 2011; Yin et al. 2012; Yoshihara et al. 2005). Liu et al. (2013) reported the identification of several flavonoid-3-*O*-glycosides in cereal species, including rice and sorghum (there is no mention of such compounds in the literature for oat). This suggests that flavonoid 3-*O*-glycosylation activity evolved from different UGT families in these species. This may explain the unusual regiospecificities reported for rice flavonoid glycosyltransferases (Ko et al. 2006; Ko et al. 2008). None of the UGT sequences identified from the transcriptomic dataset belong to groups B and M of UGTs. Those groups are conserved in all sequenced higher plants but are comprised of only a few members in each species. The restricted diversification of these groups and their conservation through evolution might suggest central roles played by these enzymes in critical mechanisms (Yonekura-Sakakibara and Hanada 2011). The present study suggests that the mechanisms in which those enzymes are involved is not taking place in root tips or is only required in certain physiological conditions.

The family UGT88 from group E is the most represented family in the oat root tip transcriptome; the other UGT families from group E both consist of a single member in oat, AsGT06751 (UGT72) and AsGT16525 (UGT71). Group E is generally the most populated group in sequenced species (Caputi et al. 2011), but in *A. thaliana* the distribution of UGTs is inverted; the UGT71 and UGT72 families are overrepresented (15 and 9 members respectively) compared to the UGT88 family (UGT88A1 being the only member of the latter). This may reflect a different functional evolution of these three families between dicot and monocot species. The

function of UGT from group E is still uncertain but dicot members of the UGT88 family have been shown to be exclusively involved in flavonoid glycosylation (Ko et al. 2008; Noguchi et al. 2009; Noguchi et al. 2007; Ogata et al. 2005; Ono et al. 2010b). The UGT72 and UGT71 families have been identified as abscisic acid and monolignol glycosyltransferases respectively in *A. thaliana* (Lanot et al. 2006; Lim et al. 2005; Priest et al. 2006; Priest et al. 2005). Very few enzymes have been characterised in group E from monocot species except an *in vitro* study with rice UGTs using flavonoid acceptors (Ko et al. 2008). Group O is particularly represented in *A. strigosa* root tips despite restricted expansion in sequenced species (Caputi et al. 2011). Functional analysis available to date suggests that UGTs from group O are involved in glycosylation of cytokinins prior to transport or storage (Martin et al. 2001; Martin et al. 1999). Expression of these enzymes in *A. strigosa* roots is in accordance with a potential role in cytokinin transport, considering that adenosine-type cytokinins like zeatin are generally synthesized in the roots (Bajguz and Piotrowska 2009); *cis*-zeatin-*O*-glucosyltransferase isolated from maize (Martin et al. 2001) is expressed primarily in roots, as AsGT05827 from oat (Fig. 3.5). *A. strigosa* roots contain two intriguing UGTs (AsGT11i11 and AsGT3i21) that have low similarities with characterised UGTs; the closest enzyme to these within the phylogenetic tree is the *C*-glycosyltransferase of flavonoids from rice (Brazier-Hicks et al. 2009). Flavonoid-*C*-glycosides have been reported from oat, and AsGT11i11 and AsGT3i21 may be involved in their production (Soriano et al. 2004).

Particular emphasis was placed on group D (UGT73) enzymes since functional studies suggest that several members of this group have evolved activities in saponin glycosylation in other species. UGT73s have been reported to have activity towards triterpenoids (Achnine et al. 2005; Augustin et al. 2012; Naoumkina et al. 2010; Sayama et al. 2012; Shibuya et al. 2010) or related compounds (e.g. phytosterols, glycoalkaloids, brassinosteroids) (Husar et al. 2011; Itkin et al. 2013; Itkin et al. 2011; Kohara et al. 2005; Kohara et al. 2007; Moehs et al. 1997; Poppenberger et al. 2005); flavonoid glycosyltransferase activities are also preponderant in group D (Hirotani et al. 2000; Jones et al. 2003; Kim et al. 2006; Ko et al. 2006). Interestingly, group D of soybean has extended considerably compared to other dicot species (Caputi et al. 2011); this might be related to the large number of saponins produced by this species (37 saponins identified) (Huhman and Sumner

2002). Phylogenetic analysis comparing the entire UGT73 families of *A. thaliana* and *O. sativa* suggests differential evolution of monocots and dicot within this family. UGT73s from dicot species have expanded in two major branches (D1 and D2) while monocot UGT73s have diversified further and formed 5 monophyletic branches (M1 to M5, Fig. 3.3). UGT73s from monocots remain largely uncharacterised except for RUGT-5 a flavonoid-3-*O*-glucosyltransferase from rice (Ko et al. 2006). *A. strigosa* enzymes are distributed across four of the branches (M1, M2, M3 and M5) identified for monocot UGT73s. Functional analysis of *A. strigosa* enzymes should increase our understanding of the evolution of monocot UGT73s.

3.3.3   Insights into synthesis of oat root glycosides

To my knowledge, no extensive metabolomic analysis has been performed on oat roots other than our own work on triterpene analysis. Nevertheless several glycosides have been identified from oat root tissues. Flavonoid-*C*-glycosides have been reported from oat and may be involved in defence mechanism against nematodes (Soriano et al. 2004). A flavonoid-*C*-glycosyltransferase has been reported from rice (Brazier-Hicks et al. 2009). This enzyme has not been included in the phylogenetic tree because of the impact of its unusual primary structure on the alignment; it shows only low identity with the oat UGTs identified. The antimicrobial coumarin scopoletin-7-*O*-glucoside (scopolin) is exuded from oat root (Fay and Duke 1977). Scopoletin-7-*O*-glucosyltransferase activity was reported for *N. tabacum* enzyme NtGT1a (Taguchi et al. 2001); this enzyme has broad acceptor specificity *in vitro* but gives enhanced resistance against viral infection when over-expressed in tobacco (Gachon et al. 2004; Matros and Mock 2004). NtGT1a is part of family UGT71 but displays only 38% identity with AsGT16525 from oat. Blumenin is a diglycoside of blumenol, a C-13 terpenoid that originates from the carotenoid pathway. This compound accumulates in response to arbuscular mycorrhizal colonisation (Maier et al. 1995). No glycosyltransferases activities have been reported toward blumenol, and it is therefore difficult to speculate about potential UGT families involved in blumenin biosynthesis.

Unfortunately, there are no clear candidate glycosyltransferases for these three glycosides reported in oat roots. Considering the fact that those chemicals are part of defence mechanisms (e.g. flavonoid-*C*-glycosides or scopolin) or secreted during symbiosis (e.g. blumenin) it is not surprising that their biosynthetic genes are poorly expressed in artificial laboratory conditions.

3.3.4   Selection of UGTs potentially involved in triterpenoid glycosylation

The analysis of the oat UGT collection has identified several candidate avenacin glycosyltransferases that can now be subjected to functional analysis (Fig. 3.7). On the basis of the phylogenetic analysis, the entire UGT73 family from oat was selected based on the known contribution of this family of enzymes to triterpenoid glycosylation in other plant species. Sterol glycosyltransferases from family UGT80 were also included in our selection due to the implication of sterol-derived enzymes in avenacin biosynthesis (e.g. SAD1 and SAD2). AsGT3i21 and AsGT11i11 are potentially interesting due to their distant relationship to any characterised glycosyltransferase. Additionally, protein accumulation in root tips and gene expression profiling of AsGT11i11 corroborates with SAD10. AsGT21p16 and AsGT25n16 were selected due to their *Sad* gene-like expression profiles. UGTs that were abundant in the root tips compare to the root elongation zone based on orbitrap analysis were also selected for further functional analysis. These were AsGT16525, AsGT10326, AsGT05827, AsGT23586A and AsGT23586B. In total 19 UGTs were selected for cloning for functional analysis. A summary of these is given in table 3.4.
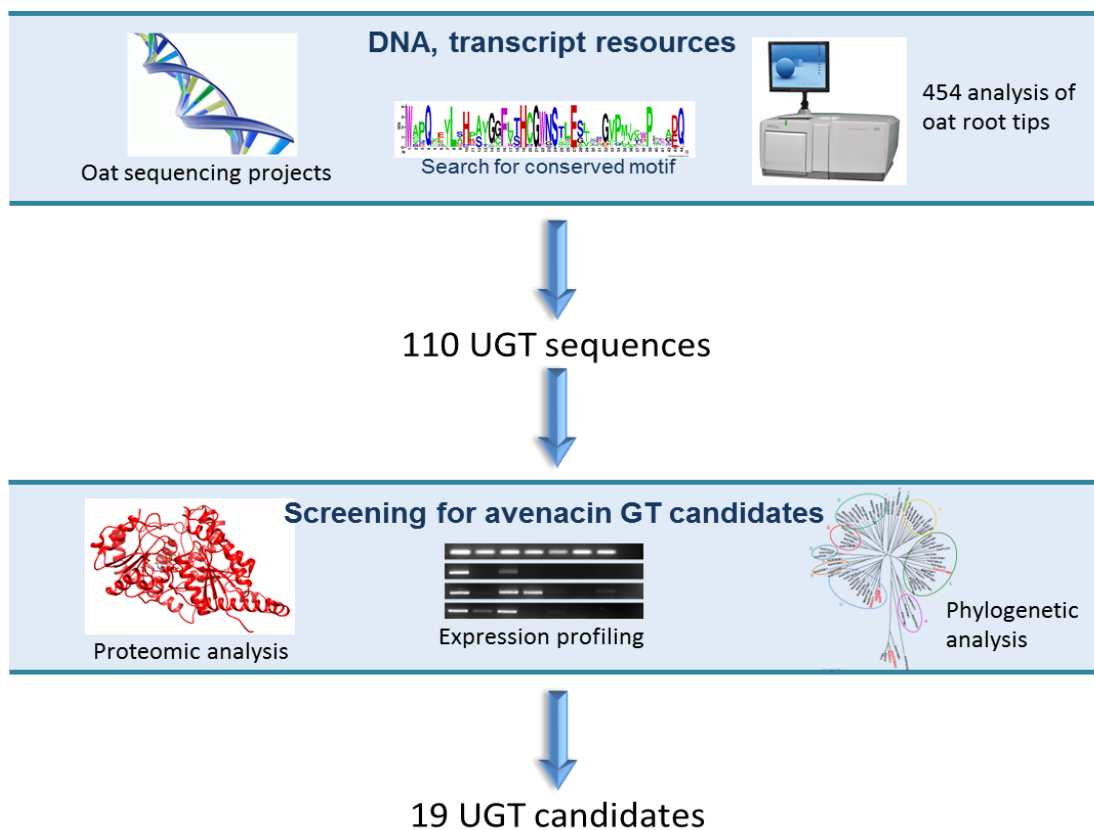
Figure 3.7. Summary of the strategy used to delineate a restricted number of avenacin glycosyltransferases candidates.

Table 3.4: Summary table of selected avenacin glycosyltransferases candidates. Protein sequences of those UGTs are displayed in supplementary data S.4.

A.strigosa UGTs	UGT family	Proteomic analysis	Expression profiles							Closest UGT with known activity
			RT	EZ	YR	YL	ML	St	F	
SAD10	UGT74	RT+	+++	-	++	-	-	-	-	N-methylanthranilate GlcT (SAD10)
AsGT03i21	nd	ND	+	++	++	-	-	-	-	Flavonoid-*O*-glucosyltransferase
AsGT11i11	nd	RT+	+++	-	++	-	-	-	-	Flavonoid-*O*-glucosyltransferase
AsGT14h20	UGT73	ND	++	-	+	-	-	-	-	Brassinosteroid 23-*O*-glucosyltransferase
AsGT14h21	UGT73	ND	+++	-	+	+	-	-	-	Hederafgenin 3-*O*-glucosyltransferase
AsGT16f23	UGT73	ND	++	+	++	++	+	+	+	Hederagenin 3-*O*-glucosyltransferase
AsGT27f7	UGT73	ND	+++	++	++	++	+	-	+	Brassinosteroid 23-*O*-glucosyltransferase
AsGT02436	UGT73	ND	+++	+	++	-	-	-	-	Brassinosteroid 23-*O*-glucosyltransferase
AsGT24i2	UGT73	ND	+++	++	++	+++	++	++	+++	Flavonoid-*O*-glucosyltransferase
AsGT23781	UGT73	RT+	++	-	++	-	-	-	++	Hederagenin 3-*O*-glucosyltransferase
AsGT25n16	UGT88	ND	+++	-	++	-	-	-	-	Flavonoid 3-*O*-glucosyltransferase
AsGT21p16	UGT85	RT/EZ	+++	+	++	-	-	-	-	Iridoid glucosyltransferase
AsGT16525	UGT71	RT+	+++	++	+++	++	++	-	++	Flavonol 3-*O*-glucosyltransferase
AsGT05827	group O	RT+	+++	+	+++	+++	+	-	++	Cis-zeatin-*O*-glucosyltransferase
AsGT23586A	UGT88	RT+	+++	++	+++	-	+	-	-	Flavonoid 3(7)-*O*-glucosyltransferase
AsGT23586B	UGT88	RT+	+++	+	+++	+	+	-	+	Flavonoid 3(7)-*O*-glucosyltransferase
AsGT03999	UGT80	ND	+++	+	++	++	++	+	+	Sterol glycosyltransferase
AsGT24525	UGT80	ND	+++	+	++	+	+	-	+	Sterol glycosyltransferase
AsGT24138	UGT80	ND	++	+	++	++	++	-	++	Sterol glycosyltransferase

Chapter 4 - Establishing platforms for the functional analysis of oat UGTs

4.1 Introduction

The multiple potential acceptors employed by UGTs *in planta* and the promiscuity displayed by these enzymes *in vitro* tend to make functional analysis a complex task (Vogt and Jones 2000). The magnitude of the task is reinforced by the difficulty to predict likely function of plant UGTs based on sequence similarities only. This is due to lineage specific evolution of UGT substrate specificity (Caputi et al. 2011; Yonekura-Sakakibara and Hanada 2011) and the poor understanding of acceptor recognition and reaction mechanisms (Osmani et al. 2009). On the other hand, the structural basis of sugar specificity, which is beginning to be understood, emphasises that UGT are generally very selective for their sugar donor (Kubo et al. 2004; Noguchi et al. 2009; Osmani et al. 2009). Therefore, innovative approaches are required to address the essential question of donor and acceptor specificity, which is a prerequisite for performing functional enzymatic assays.

The intrinsic nature of tailoring enzymes, such as UGTs, requires the use of complex modified chemicals scaffolds as substrates. Such molecules are generally not readily available (e.g. commercially) and synthetic chemistry is often unable to produce these structures in a straightforward manner. Therefore, alternative approaches are needed to investigate the mechanistic basis of the sophisticated glycosylation patterns achieved by UGTs with specialised plant metabolites. In this chapter, new platforms have been developed and used to investigate catalytic properties of *A. strigosa* UGTs.

4.1.1   Expression and purification of recombinant UGTs

Numerous examples of functional characterisation of plant UGTs have relied on *in vitro* enzymatic analysis of recombinant enzymes (Kannangara et al. 2011; Mohamed et al. 2011; Nagatoshi et al. 2012). *Escherichia coli* has been used successfully for expression of many active recombinant UGTs from plants. For example, Lim et al. (2003) cloned 110 UGTs from *A. thaliana*; expression of these UGTs in *E. coli* enabled a large-scale analysis of the catalytic properties of the recombinant enzymes towards hydroxycoumarins and benzoates (Lim et al. 2003; Lim et al. 2002). Various *E. coli* strains have been used to express plant enzymes. Of these the BL21 Rosetta strain is particularly useful for expression of eukaryotic proteins because it contains the pRARE plasmid, which supplies tRNAs for codons that are rarely encountered in prokaryotes (Novy et al. 2001).

Soluble histidine-tagged enzymes are convenient to purify using immobilized metal ion affinity chromatography (IMAC). The properties of polyhistidine-tagged proteins are generally unaffected compared to their wild type counterparts (Terpe 2003). Exceptionally, formation of dimers or alterations of catalytic properties can occur (Halliwell et al. 2001; Wu and Filutowicz 1999). *N*-Terminal tags are generally used for purification of plant UGTs, and no difficulties have been reported concerning functional analysis of recombinant *N*-terminal polyhistidine-tagged UGTs (Hansen et al. 2009; Itkin et al. 2011; Kannangara et al. 2011). Consistent with this, the synthesis of 6-deoxy-6-fluoro-β-D-glucosyl *N*-methylanthranilate from UDP-6-deoxy-6-fluoro-α-D-glucose and *N*-methylanthranilate was carried out using purified recombinant *N*-terminal 9xhistidine-tagged SAD10 (Caputi et al. 2013).

4.1.2   Functional analysis of recombinant UGTs

Despite their importance in plant metabolism and the growing number of genes that are predicted to encode UGTs from large-scale genome and transcriptome sequencing projects, functional analysis of UGTs remains challenging (Caputi et al. 2011). UGTs have functions in the later steps of many metabolic and biosynthetic pathways. The natural substrates of these enzymes are therefore often structurally complex molecules. In the case of avenacin A-1, the likely biosynthetic model implies the addition of three hydroxyl groups, an epoxide and an aldehyde group prior to glycosylation of triterpene scaffold β-amyrin. Consequently appropriate glycosyl acceptors for UGTs are difficult, if not impossible, to access commercially or through chemical synthesis. Enzymatic or acidic hydrolysis of the glycosylated natural product may offer an alternative route to generate the necessary acceptor molecule for use in functional assays (Ikeda et al. 1998; Osbourn et al. 1995), but again this is far from trivial. Indeed, the epoxide functional group of avenacin between C-12 and C-13 is unstable in acidic conditions used in deglycosylation reactions (Geisler et al. 2013; Ikeda et al. 1998). The enzymatic approach is complicated by the presence of two sugar units in the composition of the trissacharide, β-D-glucose and α-L-arabinose (Conchie et al. 1968; Crombie et al. 1984).

In vitro assays have previously been used successfully to identify likely saponin glycosyltransferases from various plant species. Available triterpene scaffolds, including β-amyrin and derivatives, oleanolic acid, hederagenin and medicagenic acid have been used for investigation of enzymes that generate monoglycosylated triterpenes (Achnine et al. 2005; Augustin et al. 2012; Meesapyodsuk et al. 2007; Naoumkina et al. 2010). A range of different strategies have been used to overcome the problems associated with the restricted number of triterpenoids acceptors available. For example, Sayama et al. (2012) exploited natural variation in the triterpene glycoside saponin content in soybean to extract an intermediate of soyasapogenol A lacking a sugar residue. They then used the purified compound as a glycosyl acceptor for enzymatic assays. Chemical deglycosylation of saponins has also been commonly used to generate the required glycosyl acceptors (Augustin et al. 2012; Ikeda et al. 1998; Shibuya et al. 2010).

4.1.3 Importance of triterpenoid glycosyltransferase characterisation

The bioactivity of saponins is due to their amphiphilic properties, allowing them to complex with lipid bilayers of cell membranes whilst altering their integrity (Augustin et al. 2011). The glycosidic moiety forms the hydrophilic part of saponins, where a triterpenoid constitute the apolar part of the molecule. Therefore, the glycosidic moiety is essential to saponin bioactivity; Armah et al. (1999) showed that incomplete glycosylation abolished the permeabilization effect of avenacin towards lipid bilayers. Therefore enzymatic processes leading to the synthesis of avenacin trisaccharide need to be understood in order to engineer the anti-fungal properties attributable to avenacin into susceptible crops species.

Functional characterisation of saponin glycosyltransferases presents numerous challenges. The promiscuity shown by many plant UGTs toward their acceptors tends to complicate interpretation of *in vitro* studies. Indeed, activity displayed by a recombinant enzyme *in vitro* may be unrelated to its physiological function. For example, recombinant UGT71G1 shows a preference for flavonoid acceptors *in vitro*, despite a role in triterpenoid glycosylation *in planta* [corroborated by mutant plant lines affected in saponin content] (Achnine et al. 2005; He et al. 2006; Shao et al. 2005). *In vitro* studies on saponin glycosylation are also challenging due to the limited availability of appropriate substrates. The only avenacin precursor available commercially is β-amyrin, but the physicochemical properties of this compound are dramatically different to those of deglycosylated desacyl avenacin (Fig. 1.11), the hypothetical natural acceptor substrate for avenacin glycosyltransferases.

4.1.4 Aims

In this chapter, the catalytic properties of selected glycosyltransferases were investigated using recombinant *A. strigosa* UGTs *in vitro*. Recombinant enzymes were expressed in *E. coli* and purified for *in vitro* assays of function. Sugar donor specificity was investigated using an innovative enzymatic assay based on UGTs general acceptor 2,4,5-trichlorophenol and enzyme activity was tested toward potential triterpenoid acceptors.
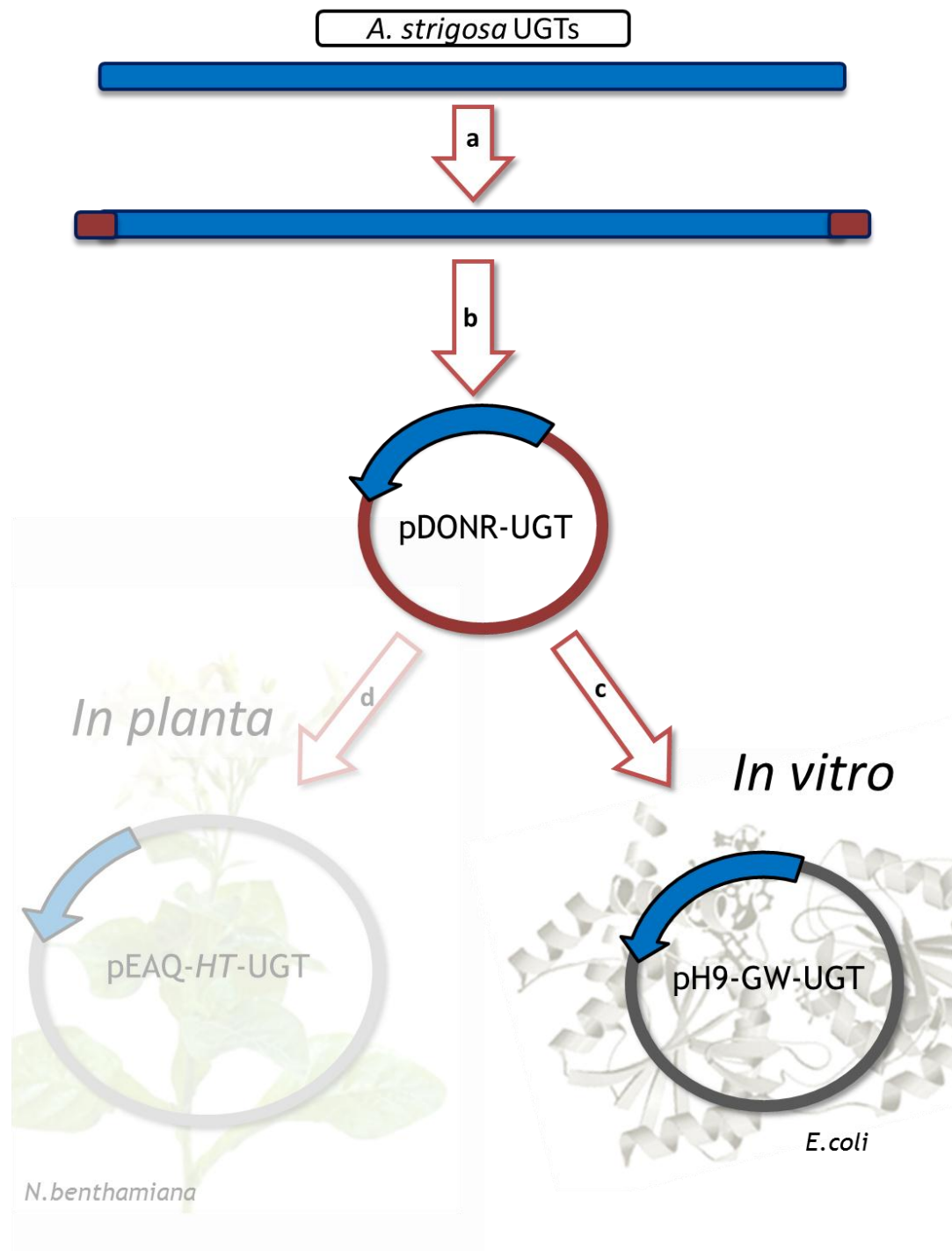
Figure 4.1: Overview of the strategy for cloning *A. strigosa* UGT coding sequences into *E. coli* expression vectors. A two-step PCR protocol was used to amplify the UGT coding sequences and add terminal AttBs adapters (a). The UGT coding sequences were then inserted into the GATEWAY entry vector pDONR207 using the BP clonase reaction (b) (Hartley et al, 2000). Two expression vectors were constructed from each pDONR construct using LR clonase reactions (Hartley et al, 2000). The *E.coli* expression vector, pH9-GW, a GATEWAY-compatible variant of pET-28 with an N-terminal 9xHis tag was used for expression of recombinant histidine-tagged enzymes for *in vitro* assays of function (c).

4.2 Results and discussion

4.2.1 Cloning strategy for oat triterpene glycosyltransferase candidates

Previous work (Chapter 3) led to the identification of a suite of *A. strigosa* UGTs for evaluation as avenacin glycosyltransferase candidates (table 3.4). The coding sequences of these UGTs were cloned into appropriate expression vectors to enable functional characterisation.

A GATEWAY cloning strategy was developed to express the *A. strigosa* UGT candidate sequences in selected heterologous systems (Fig. 4.1). The GATEWAY recombination cloning technology allows easy transfer of the UGT coding DNA sequences (CDSs) from one expression vector to another via a site-specific recombination process catalysed by clonases (Hartley et al. 2000). This system is particularly convenient for the propagation of an insert in various vectors to create constructs for multiple purposes. The strategy used here relies on the expression of recombinant *A. strigosa* UGTs in *E. coli* to perform *in vitro* assays and in *N. benthamiana* for co-expression with other triterpene biosynthetic enzymes *in planta* (Chapter 5). The selected UGT sequences were inserted into pDONR207 to obtain entry clones (Fig. 4.2). Recombination using LR clonase then enabled the insertion of the UGT CDSs into appropriate expression vectors (Fig. 4.3). The pEAQ-*HT*-Dest1 vector is a GATEWAY-compatible version of the pEAQ vector series (Sainsbury et al. 2009). The pH9-GW is an expression vector designed to produce N-terminal 9 histidine-tagged recombinant protein in *E. coli*.
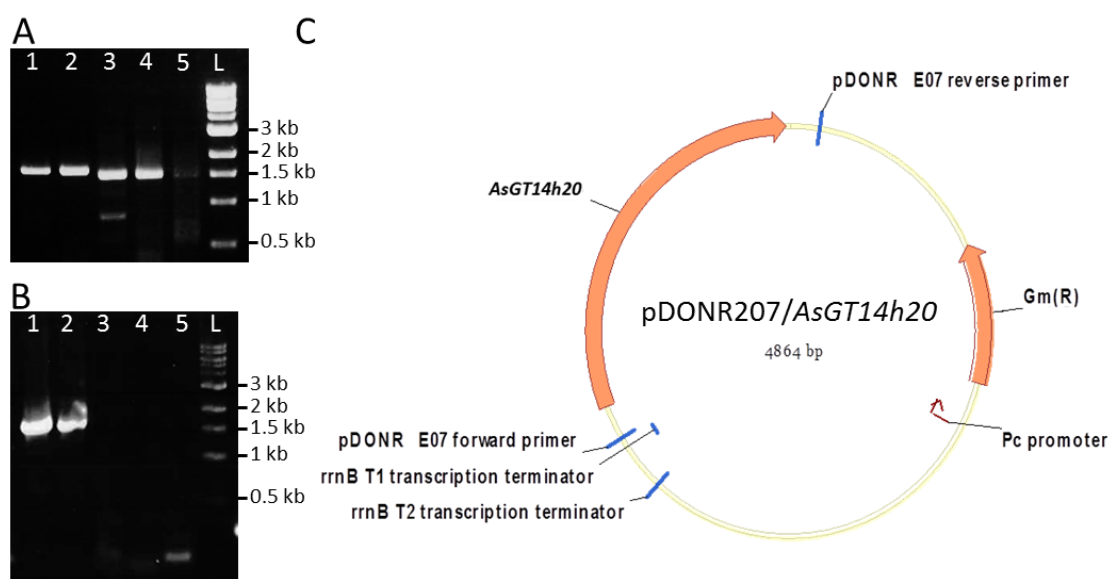
Figure 4.2: Cloning of *AsGT14h20* coding sequence into pDonr207. A) DNA gel of *AsUgt* CDS amplified from *A strigosa* root tip cDNA with appropriate AttBs adapters. Lane 1, *AsGT14h20*; Lane 2, *AsGT24i2*; Lane 3, *AsGT27f7*; Lane 4, *AsGT25n16*; Lane 5, *AsGT05827*; Lane L, 1 kb DNA Ladder (New England Biolabs®). *AsGT14h20*, *AsGT24i2* and *AsGT27f7* were specifically amplified under the PCR conditions used**. B).** Amplification of five clones obtained following the BP clonase reaction for the amplification product of *AsGT14h20* and pDONR207 (Lanes 1-5). Lane L, 1 kb DNA Ladder (New England Biolabs®). Clones 1 and 2 have the expected size of 1503 bp and were sequenced for confirmation (see section 2.2.6). **C).** Map of the resulting pDONR construct containing the *AsGT14h20* coding sequence.

Specific primers were designed for each UGT gene selected (Section 2.1.3) and amplification conditions were optimised to obtain a single product. A two-step amplification PCR protocol was used to attach AttB1 and AttB2 adapters to the 5' and 3' ends of the UGT CDSs respectively (method in section 2.2.9). An example of amplification is given in figure 4.2.A, under PCR condition tested (55°C of annealing temperature) amplification of the expected size (see table 4.1) was obtained for *AsGT14h20*, *AsGT24i2*, *AsGT27f7* and *AsGT25n16* (lanes 1-4). *AsGT05827* transcript (lane 5) was not amplified under these conditions and additional PCR experiments were conducted in order to obtain the expected product (gradient PCR, nested PCR, modification buffer composition). Appropriate conditions for the amplification of the following UGT CDSs were not found: AsGT23781, AsGT16525, AsGT10326 AsGT24138 and AsGT24525. For each UGT sequence, colony PCR was carried out with five to eight isolated colonies that were resistant to gentamicin (method in section 2.2.12). In figure 4.2.B, two out of the five positives colonies selected for AsGT14h20 entry clones have the expected

insert (lanes 1 and 2). The entry clones incorporating a DNA fragment of the expected size (approximately 1.5 kbp) were sequenced to verify the identity of the clone (method in section 2.2.10 and 2.2.6). Only pDonr207-UGTs without mutations in the inserted CDS were used for further work. Sequencing of the entry clones obtained after amplification of contig23586 revealed the insertion of two homologous UGT sequences, 23586A and 23586B. It transpired that contig23586 was the result of an incorrect assembly of reads. The first 530 bp at the 5' end of contig23586 correspond to AsGT23586A, and the 3' end corresponds to AsGT23586B. The difficulties encountered for the cloning of the 5 CDSs mentioned above may be due to similar assembly errors. The entry clones constructed are all listed in table 4.1. The structure of the pDONR207-AsGT14h20 construct is shown as an example (Fig. 4.2.C). The pDONR207 was used primarily due to its gentamycin resistance gene being compatible with the kanamycin resistance provided by each of the two expression vectors. Additionally the *SAD10, UGT73C10* and *UGT78D3* coding sequences were cloned to serve as controls for functional analysis. As indicated in figure 1.9, SAD10 is an *A. strigosa N*-methyl anthranilate glucosyltransferase involved in avenacin biosynthesis (Owatworakit et al. 2012); to date, UGT78D3 is the only plant UGT with arabinosylation activity that has been characterised (Yonekura-Sakakibara et al. 2008); and UGT73C10 is a triterpene 3-*O*-glucosyltransferase that is able to glucosylate oleanane-type triterpenes from *Barbarea vulgaris* (Augustin et al. 2012).

The corresponding expression vectors (pH9-GW-UGT and pEAQ-*HT*-UGT) were obtained using the LR clonase reaction (Fig. 4.3). Cloned DNA fragments inserted within these expression vectors were amplified by PCR and a single colony incorporating an insert of the expected size was used for further studies. Expression vectors were initially replicated in *E. coli* DH5α and then transferred into either *A. tumefaciens* strain LBA4404 (for pEAQ-*HT*-UGT vectors) or *E. coli* strain BL21 (for pH9-GW-UGT vectors). Transformation methods are described in section 2.2.10.
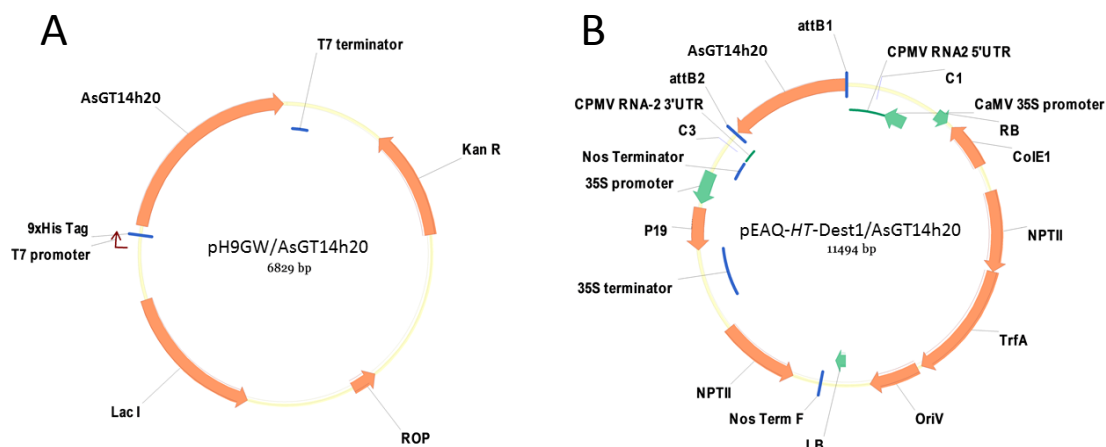
Figure 4.3: Maps of expression vectors containing the *AsGT14h20* CDS insert with key features labelled and total sizes indicated. A). Map of pH9-GW-AsGT14h20. Expression of recombinant *N*-terminal tagged *AsGT14h20* is under the control of the LacI promoter and is induced by isopropyl β-D-1-thiogalactopyranoside (IPTG). ROP, repressor of primer gene; Kan R, kanamycin resistance gene**. B).** Map of pEAQ-HT-Dest1:AsGT14h20. Heterologous expression of *AsGT14h20* in *N. benthamiana* is under the control of the constitutive 35S promoter. P19 is a suppressor of silencing; LB and RB are the T-DNA left and right borders respectively; OriV is the origin of replication of pRK2; ColEI is the replication origin of pBR322; NPTII, neomycin phosphotransferase; TrfA is a replication-essential locus.

4.2.2 Characterisation of recombinant oat UGTs following *E. coli* expression

Production and purification of recombinant enzymes was a prerequisite for the functional analysis of *A. strigosa* UGTs *in vitro*. A series of experiments was performed in order to improve the expression of soluble enzymes in *E. coli* and to purify the recombinant UGTs.

E. coli expression vectors (pH9-GW-UGTs), as described above, were used to transform *E. coli* strain BL21 (DE3) (method in section 2.2.10). Preliminary expression experiments were performed with the following induction conditions: addition of 0.1 mM IPTG to cells at O.D.$_{600nm}$ 0.6, and subsequent harvesting of the cells after overnight culture at 18°C. These are common conditions used for the expression of many plant UGTs from *E. coli* transformants (Frydman et al. 2013; Landmann et al. 2007; Owatworakit et al. 2012; Yoon et al. 2012). Separation and analysis of the cell lysate insoluble and soluble fractions suggested that most of the recombinant UGTs were expressed under the conditions tested (Fig. 4.4). Accumulation of new proteins with the expected size (50-55 kDa, details in table

4.1) is seen in most of the insoluble fractions. Recombinant UGTs were not detected in the soluble fraction for most of the constructs tested here, suggesting a large portion of these enzymes has been expressed as inclusion bodies. Despite a substantial amount of protein accumulated in the insoluble fraction for AsGT16f23, AsGT16h6, AsGT24i2, AsGT25n16, AsGT23586A and AsGT23586B the protein is hardly visible in their respective soluble fractions. Lower levels of expression are detected exclusively in the insoluble fraction for SAD10, AsGT3i21, AsGT14h20, AsGT27a12 and AsGT005827. AsGT11i11 does not seem to be expressed under conditions tested. AsGT14h21, AsGT21p16, AsGT27f7, AsGT02436 and AsGT03999 were cloned after the UGT mentioned above and consequently were not included in these preliminary expression experiments.
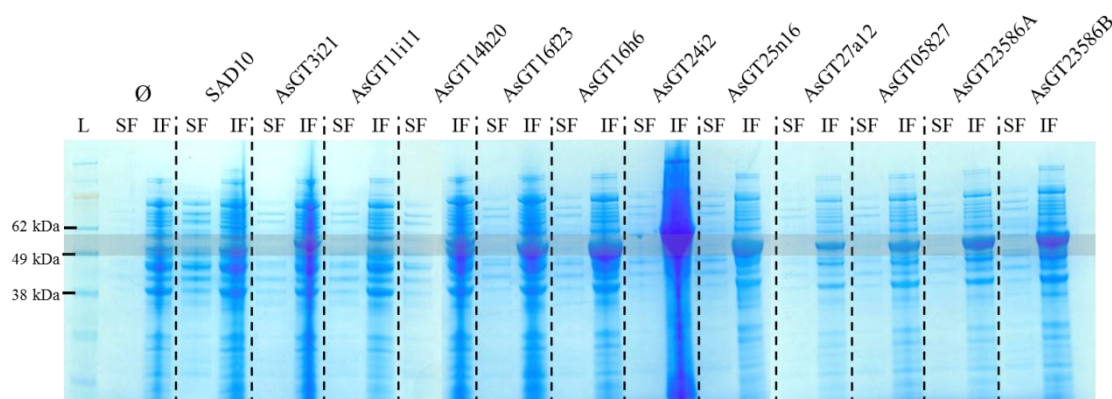
Figure 4.4: Analysis of insoluble and soluble fractions of *E. coli* BL21 transformants expressing *A. strigosa* UGTs. Coomassie-stained SDS-PAGE gel of the cell lysate of *E. coli* BL21 transformed with pH9-GW-UGT plasmids. First lane is loaded with SeeBlue® Plus2 Pre-Stained Standard, following lanes are loaded alternatively with the soluble fraction (SF) or the insoluble fraction (IF) obtained after centrifugation of the cell lysate at 13,000 rpm for 15 min. For each transformant: an equal volume of 15 µl of the soluble fraction was loaded; additionally, 2 µl of the insoluble fraction resuspended and boiled in 50 µl of dH_2O was loaded. Expected migration zone of *A. strigosa* UGTs has been highlighted in grey (50-56 kDa).

Insoluble recombinant proteins may be due to intrinsic insolubility of the enzyme (transmembrane e domain, hydrophobic regions, protein associated with membrane-bound proteins), or formation of inclusion bodies (Baneyx 1999; Baneyx and Mujacic 2004). Plant UGTs are known to be cytosolic enzymes (Bowles et al. 2006) and none of the *A. strigosa* UGTs has been predicted to have a transmembrane domain (http://prosite.expasy.org/). This may suggest insolubility is due to inclusion body formation, possibly as a result of the aggregation of misfolded proteins overexpressed in the host cell (Baneyx and Mujacic 2004).

In order to evaluate the solubility of all of the enzymes, further expression experiments were carried out using two different induction conditions (21°C for 5h, or 18°C overnight). A non-ionic detergent was added to the lysis buffer to improve solubility the yield of soluble protein (0.1 % Tween 20). Western blot analysis was performed using anti-polyHistidine antibody to allow the detection of low quantities of recombinant UGTs in the soluble fraction of the cell lysate. While AsGT14h20, AsGT24i2, AsGT02436, AsGT05827, AsGT23586A and AsGT23586B were detectable using 21°C for 5h (Fig. 4.5.A) more UGTs were detectable following induction at 18°C overnight (Fig. 4.5.B). Only AsGT24i2, AsGT05827, AsGT23586Aand AsGT23586B were readily detectable on Ponceau-stained membranes suggesting a higher expression of these recombinant proteins. All of the UGTs were detectable in the soluble fraction under at least one condition used except for AsGT27f7, AsGT27a12 and AsGT03999. All of the recombinant enzymes are not represented in figure 4.5 since some of them were cloned later and processed independently (e.g. AsGT27f7, AsGT03999, AsGT21p16 AsGT14h21 and AsGT02436; data not shown). A very low level of recombinant protein was detected for AsGT3i21, AsGT11i11 and AsGT25n16 (Fig. 4.5).
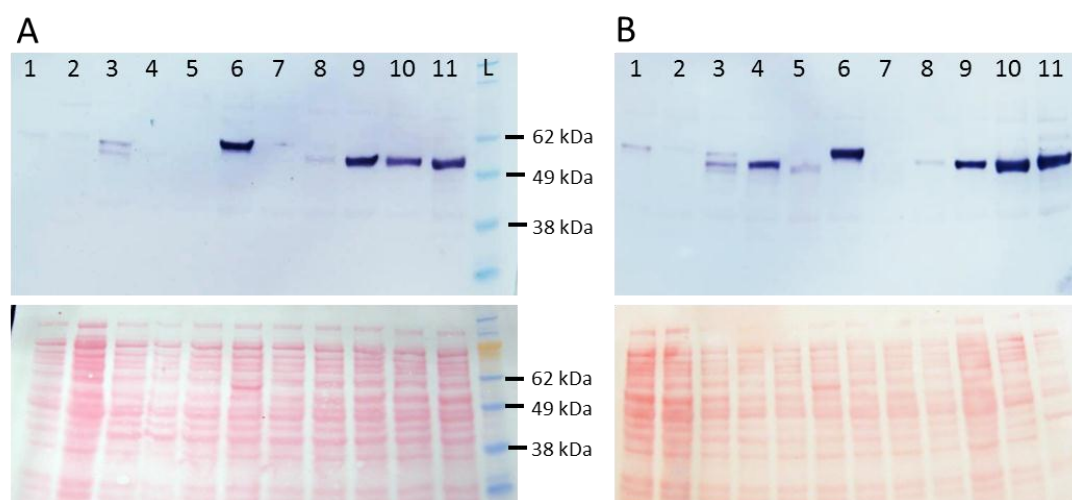
Figure 4.5: Western blot of soluble recombinant UGTs from *A. strigosa*. Western blot analysis was carried out using the soluble fraction of cell lysates from *E. coli* BL21 transformants expressing *A. strigosa* recombinant UGTs. Colorimetric detection of recombinant His-tagged UGTs was performed using monoclonal anti-polyHistidine antibody (Sigma). Lane 1, AsGT3i21; lane 2, AsGT11i11, lane 3, AsGT14h20, lane 4, AsGT16f23, lane 5, AsGT16h6, lane 6, AsGT24i2, lane 7, AsGT25n16, lane 8, AsGT27a12, lane 9, AsGT05827, lane 10, AsGT23586A; lane 11, AsGT23586B and lane L, SeeBlue® Plus2 Pre-Stained Standard. **A.** Analysis after 5 hours of induction at 21°C. **B.** Analysis after overnight induction at 18°C. Ponceau-stained membranes are displayed in the lower panel.

To improve the yield of our soluble recombinant UGTs, a suite of different *E. coli* strains and induction conditions was evaluated. The expression vectors (pH9-GW-UGT) were used to transform *E. coli* strains BL21 Rosetta (Rosetta 2 (DE3); Novagen) and Lemo21 (Lemo21 (DE3); NewEngland BioLabs). These bacterial strains are expected to aid the expression of eukaryotic proteins (Rosetta) and modulate protein expression to increase solubility (Lemo21) (Novy et al. 2001; Wagner et al. 2008). Addition of glycerol (5%) into the media and induction at 21°C or 18°C were also tested. Lower temperature of induction might help reducing the formation of inclusion bodies and favour correct protein folding (Baneyx 1999). Culture media supplemented by glycerol have been used by previous lab members to increase recombinant protein solubility. Each of the UGT clones that had been shown to produce a low amount of soluble recombinant UGT protein previously were included in experiments to systematically compare the eight possible conditions (two *E. coli* strains, two induction temperatures, +/- glycerol). Following cell lysis and isolation of the soluble fraction, histidine-tagged proteins were enriched using immobilised nickel affinity resin (method in section 2.2.13). Eluted fractions (250 mM imidazole) were analysed using Coomasie-stained SDS-PAGE.

The different conditions tested had substantial effects on the amount of recombinant enzyme in solution (Fig. 4.6, showing AsGT14h20 as an example). AsGT14h20 (54.4 kDa) is labelled on the figure, the three larger bands stained on the gel have been observed for all UGTs and empty vector controls and might be attributable to *E. coli* endogenous enzymes rich in histidine and retained by the IMAC column. Optimal expression conditions for AsGT14h20 were obtained at 18ºC after overnight culture of *E. coli* Rosetta transformant. Recombinant protein of the expected size was detected for most of the UGT clones except AsGT03999. AsGT03999 belongs to sub-family UGT80, a family that is involved in sterol glycosylation. This family of enzymes are known to be poorly soluble due to their apolar *N*-terminus, which is involved in membrane anchorage (Chaturvedi et al. 2011; Grille et al. 2010). Generally *E. coli* Rosetta strain was the best suited for expression of the recombinant UGTs, except for AsGT3i21. A higher production level of soluble AsGT3i21 in Lemo21 cells suggests that this protein might be insoluble or toxic to *E. coli*. The presence of the protein in the insoluble fraction of cell extracts and the normal growth rate of the transformant bacteria is an indication

of AsGT3i21 insolubility. The majority of the UGTs are slightly better expressed overnight at 18ºC than at 21 ºC for 5 hours, which correlates with the higher biomass produced under these conditions (data not shown).
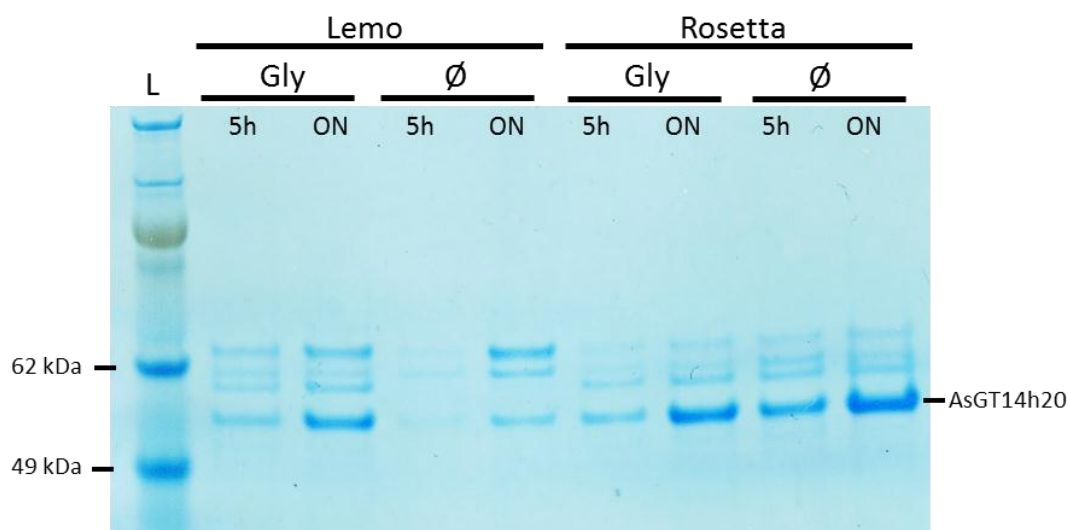
Figure 4.6: Comparison of different expression conditions for recombinant AsGT14h20. Coomassie-stained SDS-PAGE gel of the enriched fraction for recombinant AsGT14h20. Eight different expression conditions were compared: induction at 21°C for 5 hours (5h), or at 18°C overnight (ON); cell culture in LB medium (Ø) or in LB medium supplemented with 5% glycerol (Gly); Rosetta (Novagen) or Lemo21 (New England Biolabs) *E.coli* BL21 (DE3) strains were transformed with pH9-GW-AsGT14h20. Optimal expression of recombinant AsGT14h20 was obtained for overnight induction of Rosetta strain.

The optimal expression conditions established from these screens were then used to purify larger quantities of each recombinant UGT (Table 4.1). Histidine-tagged proteins from cell lysate of 500 mL cultures were purified by immobilized-metal affinity chromatography (IMAC) using an AKTA purifier system. Fractions containing the eluted recombinant enzyme were concentrated and analysed by SDS-PAGE (Fig. 4.7). Concentrations of the purified enzymes were estimated using the Bradford method (Method in section 2.2.17; table 4.1). Some of the recombinant UGTs were weakly expressed and consequently difficult to purify; in those cases the concentration of the recombinant UGT was estimated based on intensity comparison of the bands observed on Coomasie-stained gels.
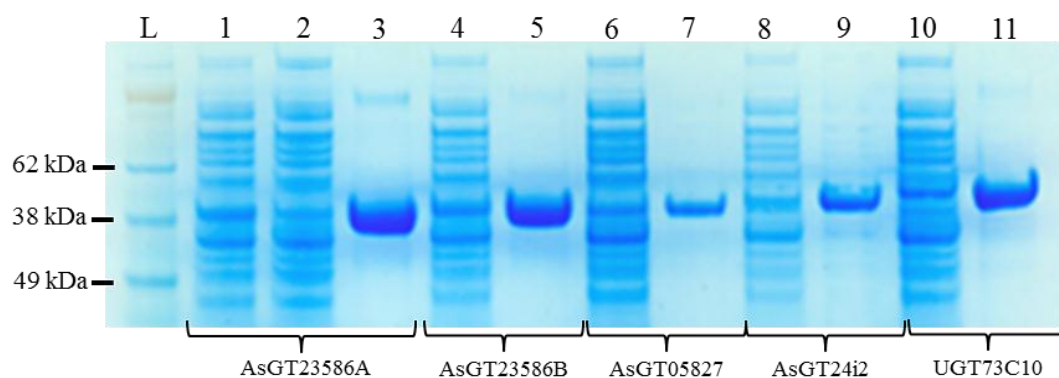
Figure 4.7: Coomassie blue-stained SDS-PAGE gel of crude soluble fractions and IMAC-purified recombinant *A. strigosa* UGTs. Lanes 1, 4, 6, 8 and 10, soluble cell lysates of induced pH9GW-UGTs transformants; lanes 3, 5, 7, 9 and 11, immobilized metal ion affinity chromatography (IMAC)-purified recombinant UGTs; L, SeeBlue® Plus2 Pre-Stained Standard. Lanes 1-3, expression and purification of AsGT23586A (lane 2 is the fraction of the soluble cell lysate unbound to the IMAC column); lanes 4-5, expression and purification of AsGT23586B, AsGT05827 (lanes 6-7), AsGT24i2 (lanes 8-9), UGT73C10 (lanes 10-11).

Table 4.1: Expression of *A. strigosa* recombinant UGTs in *E. coli*.

UGT	CDS length	Protein MW•	*E.coli* strain	Induction*	Glycerol†	pProtein yield‡
SAD10	1392 pb	50.7 kDa	BL21	21°C	-	3.2 mg/L
AsGT3i21	1530 pb	54.9 kDa	BL21 Lemo21	18°C	Yes	n.d
AsGT11i11	1530 pb	55.0 kDa	BL21 Rosetta	18°C	No	n.d
AsGT14h20	1503 pb	54.4 kDa	BL21 Rosetta	21°C	No	1.0 mg/L
AsGT14h21	1488 pb	53.3 kDa	BL21 Rosetta	18°C	No	n.d
AsGT16f23	1488 pb	55.0 kDa	BL21 Rosetta	21°C	No	0.27 mg/L
AsGT16h6	1425 pb	51.3 kDa	BL21	18°C	No	n.d
AsGT21p16	1470 pb	53.7 kDa	BL21 Rosetta	18°C	No	n.d
AsGT24i2	1575 pb	56.2 kDa	BL21	21°C	No	4.1 mg/L
AsGT25n16	1416 pb	51.1 kDa	BL21 Lemo21	18°C	Yes	n.d
AsGT27f7	1575 pb	56.2 kDa	BL21 Rosetta	18°C	No	n.d
AsGT02436	1467 pb	52.0 kDa	BL21 Rosetta	18°C	Yes	0.66 mg/L
AsGT05827	1404 pb	51.0 kDa	BL21	21°C	No	0.18 mg/L
AsGT23586A	1446 pb	52.7 kDa	BL21	21°C	No	3.5 mg/mL
AsGT23586B	1446 pb	52.8 kDa	BL21	21°C	No	1.4 mg/mL
AsGT03999	1863 pb	68.4 kDa	ND	ND	ND	n.d
UGT78D3	1377 pb	50.0 kDa	BL21 Star	24°C	No	n.d
UGT73C10	1485 pb	55.5 kDa	BL21	21°C	No	1.2 mg/mL

• MW: molecular weight of the protein in kDa.
* Induction was performed with 0.1 mM IPTG at 21/24°C for 5 hours or 18°C overnight.
† 5% of glycerol was added to the culture.
‡ The protein yields (shown as mg/L or mg/ml) were estimated using the Bradford method after purification and concentration of recombinant UGTs. n.d., protein concentrations not determined due to low levels.

4.2.3   Investigation of UGT sugar specificity

Plant UGTs use UDP-sugars as sugar donors. UDP-glucose, UDP-galactose, UDP-rhamnose, UDP-xylose, UDP-glucuronic acid and UDP-arabinose have all been reported to be used by plant UGTs of family GT1 (Bowles et al. 2006; Osmani et al. 2009). UGTs generally show a preference for a single UDP-sugar. This is clearly illustrated in the sugar specificity analysis conducted on *Lamiales* UGTs by Noguchi et al. (2009). In this study, which involved investigation of 12 UGTs using three sugar donors, the activity measured for a secondary sugar donor never exceeded 7% of that of the preferred donor (table 4.2). Therefore, it is essential to understand the sugar donor specificity of the *A. strigosa* UGTs prior to testing their activity towards potential acceptor molecules.

Table 4.2. Sugar Donor Specificity of Lamiales UGTs (Noguchi et al. 2009)

UGT	Relative Activity (%)		
	UDP-Glc	**UDP-Gal**	**UDP-GlcA**
UGT88D7	0	0	100
UGT88A7	100	0	0
UGT88D4	0	0	100
UGT88D5	0	0	100
UGT88D6	1	0	100
UGT88A1	100	3	2
UGT73A9	100	7	2
UGT73E2	100	0	0
UGT73N1	100	2	0
UGT73A7	100	6	2
UGT73A13	100	6	0

The glycosylating activity of each enzyme on three types of sugar donor (UDP-α-D-glucose, UDP-α-D-galactose, and UDP-α-D-glucuronic acid) was tested. Apigenin was used as sugar acceptor for evaluating the sugar donor specificity. Products were quantified based on peak area at A_{350}. The highest activity in the three sugar donors is set as 100%.

4.2.3.1 *2,4,5-Trichlorophenol (TCP) - A general acceptor for most UGTs*

2,4,5-Trichlorophenol (TCP) is a chlorinated molecule that is used for synthesis of herbicides (Su et al. 2012). UGTs have the ability to conjugate TCP as part of the process of detoxification of this xenobiotic (Brazier-Hicks et al. 2007b). It has been shown that many recombinant UGTs are able to glucosylate TCP as well as their natural acceptor (Messner et al. 2003). For example, in *A. thaliana* nearly 50% of GT1 UGTs are active towards TCP (Brazier-Hicks et al. 2007b), in particular UGTs from groups D, E and L (Messner et al. 2003). The broad recognition of TCP as acceptor by UGTs may be attributable to its weak acidity (with leaving group pK_a 7.4) driving the reaction towards TCP-sugar formation, as demonstrated by Gantt et al. (2011) using a subset of aryl glucosides. Here, the acceptor promiscuity of UGTs was exploited in order to carry out a comprehensive screen of potential sugar donors. The activity of each recombinant *A. strigosa* UGT was systematically tested with UDP-Glc, UDP-Gal and UDP-Ara. This analysis gave new insights into the sugar donor specificity of the *A. strigosa* UGT collection.

4.2.3.2 *Characterisation of products formed during TCP reactions*

Preliminary experiments were carried out to evaluate the feasibility of generating TCP-*O*-*β*-D-glucose (TCP-Glc) and TCP-*O*-*α*-L-arabinose (TCP-Ara) using functionally characterised glucosyltransferase SAD10 and arabinosyltransferase UGT78D3, respectively (Fig. 4.9). Preliminary experiments demonstrate that UGT78D3 used exclusively UDP-Ara as a sugar donor while SAD10 used UDP-Glc as its favoured donor and UDP-Ara to a lesser extent (Fig. 4.8; method in section 2.2.14). Formation of a product (Rt 10.0 min) was detected by HPLC after enzymatic assays in which SAD10 was incubated with UDP-*α*-D-glucose (UDP-Glc) and TCP (Fig. 4.8.A, green spectrum). The same product was found to accumulate when UGT73C10, a hederagenin 3-*O*-glucosyltransferase, was incubated with UDP-Glc and TCP (data not shown). Under similar conditions, no product was detected in the presence of UGT78D3 (Fig. 4.8.A, turquoise spectrum). A new product was formed (Rt 10.9 min) when UGT78D3 was incubated with UDP-*β*-L-arabinose (UDP-Ara) and TCP (Fig. 4.8.B, red spectrum). Incubation of SAD10 with UDP-Ara and TCP generate a small amount of product with Rt 10.9 min representing only 0.7% of the peak area obtained with UDP-Glc (Fig. 4.8.B, pink spectrum). These results suggest that the sugar specificity of SAD10 and UGT78D3

are conserved with TCP compare to their respective natural acceptors (e.g. *N*-methylanthranilate and flavonol, respectively).
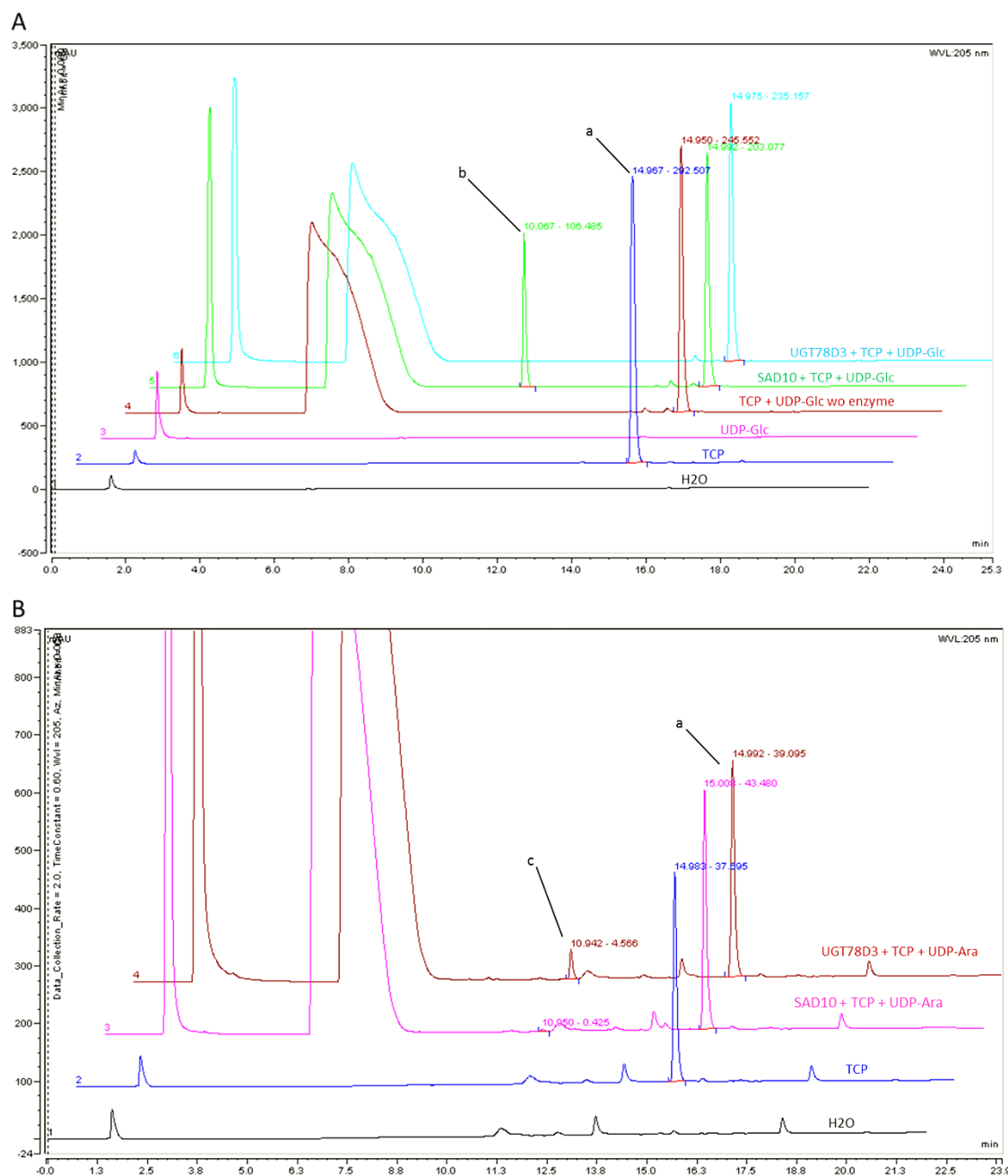
Figure 4.8: Sugar specificity of SAD10 and UGT78D3 is conserved towards TCP. A. Incubation of SAD10 and UGT78D3 with TCP and UDP-Glc. Black, blue, pink and red spectrum are controls; blank, acceptor (TCP), sugar donor (UDP-Glc) and reaction conditions without enzyme respectively. Reaction with SAD10 (green spectra) produces TCP-Glc (Rt 10.05 min), no products are detected with UGT78D3 (pale blue spectra) in similar conditions. **B.** Incubation of SAD10 and UGT78D3 with TCP and UDP-Ara. TCP-Ara (Rt 10.95 min) is produced by SAD10 and UGT78D3. Compound labelled: **a**, TCP; **b**, TCP-Glc and **c**, TCP-Ara

To confirm the accumulation of TCP-*O*-glycosides in these glycosyltransferase assays determination of the structures of the product was needed. A large-scale reaction was set up with SAD10 recombinant enzyme in the presence of UDP-Glc and TCP. The HPLC method was scaled up to isolate the product using a preparative HPLC column (method in section 2.2.14). Collection of the fraction containing the product led to isolation of 0.23 mg of white amorphous powder. The proton NMR spectrum was consistent with the identity of this compound as TCP-*O*-Glc. The spectrum is similar to that of dichlorophenol (DCP)-*O*-Glc published by Day and Saunders (2004), except for position C-5 of the phenol ring were the chlorine atom is attached on TCP (Supp. S.8). Resonances corresponding to a glucosidic moiety are visible in the 3.1 to 3.7 ppm range. The anomeric proton attached to carbon 1 of glucose (δ = 4.83 ppm) has a coupling constant (J = 7.34 Hz) characteristic of a β anomer. The protons of trichlorophenol (C-3 and C-6) appear as two singlets (δ = 7.13 ppm and δ = 7.37 ppm) instead of the doublets of HC-3 and HC-6 and the doublet of a doublet of HC-5 detected for DCP-Glc. It is consistent with the chlorine in C-5 position of TCP-Glc; consequently hydrogens of HC-3 and HC-6 are not coupling with the other hydrogen.
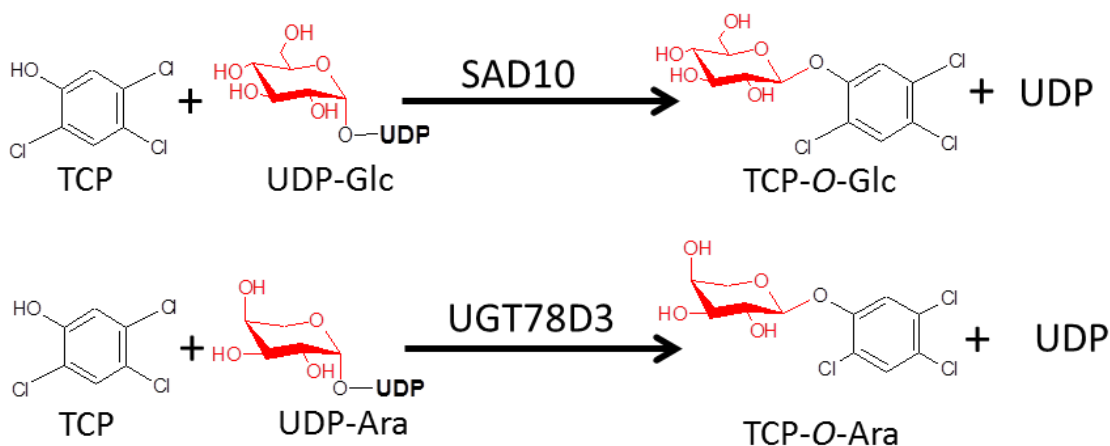
Figure 4.9: 2,4,5-Trichlorophenol glycosylation reactions catalysed by UGTs. Recombinant SAD10 catalyses the formation of 2,4,5trichlorophenol (TCP)-*O*-*β*-D-glucose by transfer of D-glucose from UDP-*α*-D-glucose onto the hydroxyl group of TCP. Recombinant UGT78D3 catalyses the formation of 2,4,5trichlorophenol (TCP)-*O*-*α*-L-arabinose by transfer of L-arabinose from UDP-*β*-L-arabinose onto the hydroxyl group of TCP.

The mass spectrometry results were in agreement with the proton NMR. A single peak was detected in full MS chromatogram (Rt 4.63 min, Fig. 4.10.A, top), confirming the purity of the product obtained by preparative HPLC. The MS spectrum extracted from the peak consists of two main ions (m/z 403, m/z 393) corresponding to the formate and chloride adducts of the TCP-*O*-glucoside (MW: 359.6 Da) (Fig. 4.10.A, middle), isotopic ions are visible for each adduct. The ion (m/z 195) is presumably TCP released as an in-source fragment. The MS2 of the formate adduct consists of a single ion (m/z 195) corresponding to a loss of formate and hexose (Fig. 4.10.A, bottom). Collectively, the proton NMR and MS data confirm the identity of the product as 2,4,5-trichlorophenol-*O*-β-D-glucopyranoside (Fig. 4.9).

Purification of TCP-*O*-Ara and TCP-*O*-Gal was not undertaken due to the significantly lower amounts of product formed for the arabinosylation and galactosylation assays. However, mass spectrometry analysis of the reaction products yielded data consistent with accumulation of TCP-*O*-Ara. The product that accumulated in the arabinosylation assays was identified as the peak eluted at 5.83 min (Fig. 4.10.B, top). It has a similar absorption spectrum to that of TCP-Glc (Fig. 4.11). The MS spectrum of the product generated by UGT78D3 in the TCP/UDP-Ara assay was composed of two major ions (m/z 375, m/z 365), which were identified as the formate and chloride adducts of TCP-*O*-Ara as they are 30 mass units lighter than TCP-*O*-Glc (Fig. 4.10.B, middle). An in-source fragmentation is also observed with release of TCP (m/z 195). The MS2 spectrum showed the expected loss of the adduct and pentose groups compared to the full ion spectrum, leading to similar MS2 spectrum than TCP-Glc (Fig. 4.10.B, bottom). Considering the similar fragmentation patterns obtained in mass spectrometry, the identical absorption spectra and the similar retention times observed for the two products in HPLC (Rt$_{TCP-Glc}$ 10.0 min and RtTCP-Ara 10.9 min) we can confidently assume that this product is 2,4,5-trichlorophenol-*O*-α-L-arabinopyranoside (Fig. 4.9).
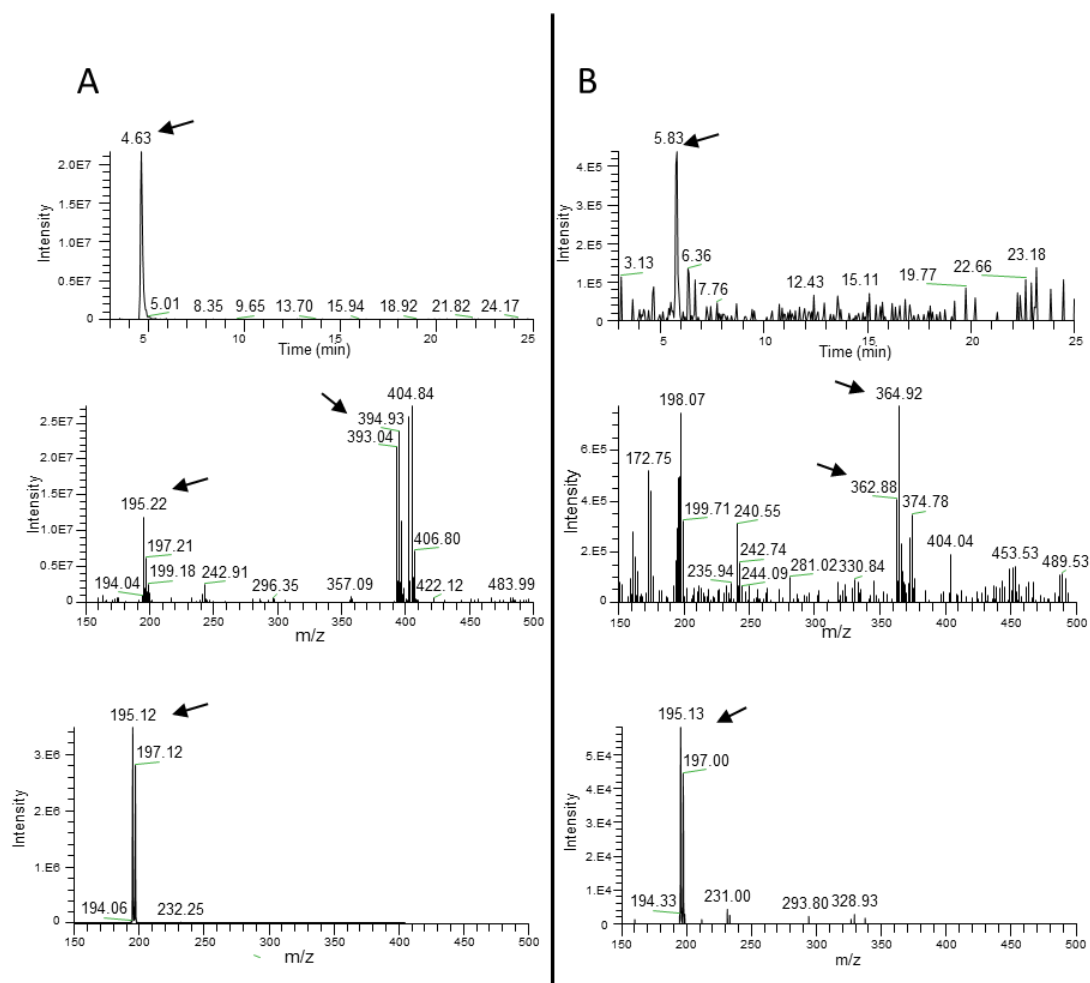
Figure 4.10: LC-MS analysis of TCP glycosides. A) LC-MS spectrum of purified 2,4,5-trichlorophenol-*O*-β-D-glucoside (TCP-Glc). Top; the EIC [M + Cl]- (m/z 393) peak at Rt 4.63 min is TCP-Glc. Middle; The extracted MS1 (middle) spectrum of the TCP-Glc peak shows chloride adduct isotopes (m/z 393 & 395), the MS2 spectrum (bottom) shows the release of TCP negative isotopic ions (m/z 195 & 197) after loss of the chloride (m/z 36) and hexose residues (m/z 162). **B).** LC-MS spectrum of the reaction products of TCP and UDP-Ara catalysed by UGT78D3. Top, the EIC [M + Cl]- (m/z 363) peak at Rt 5.83 min is TCP-Ara. The extracted MS1 spectrum (middle) of the TCP-Ara peak showscChloride adduct isotopes (m/z 363 & 365). The the MS2 spectrum (bottom) shows the release of TCP negative isotopic ions (m/z 195 & 197) after a loss of chloride (m/z 36) and pentose residue (m/z 132).

Sugar specificity analysis was primarily conducted to search for glucosyltransferase and arabinosyltransferase activities in the UGT collection in the scope of identify avenacin glycosyltransferases. Consequently mass spectrometry analysis has not been undertaken for TCP-Gal so far.
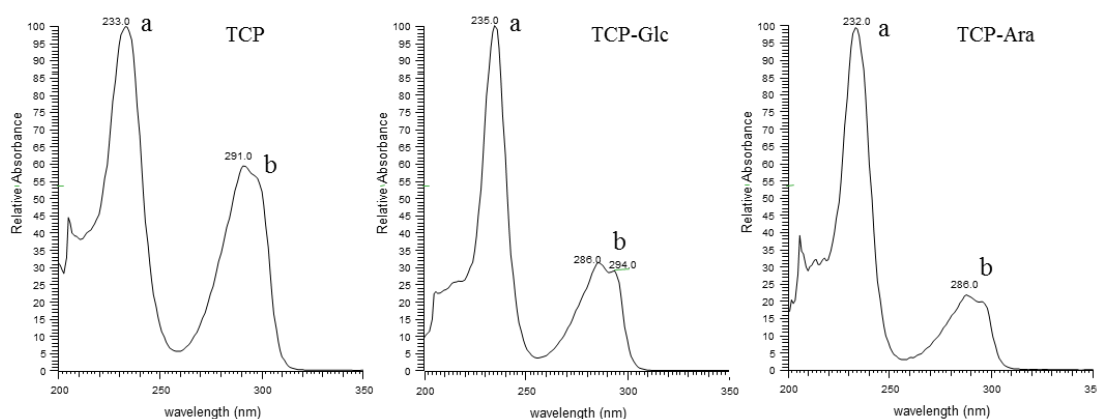
Figure 4.11: UV spectrum of TCP (left), TCP-Glc (middle) or TCP-Ara (right). TCP glycosides have similar absorbtion profiles in UV light with a maximum of absorption at 232-235nm identical to TCP (peak a). Peak b is reduced and shifted from 291 nm to 286 nm for the glycosides.

4.2.3.3   *Investigation of the activity of the oat UGT collection towards TCP*

Optimal conditions for TCP arabinosylation by recombinant UGT78D3 were obtained with substrate concentration of 0.5 mM UDP-Ara and 0.1 mM of TCP at pH 7.5 with 2 µg of recombinant UGT. TCP glycosylation reactions were performed under these conditions for each recombinant UGT with UDP-Glc, UDP-Gal and UDP-Ara (method in section 2.2.14). We based our analysis on the reasonable supposition that TCP glycosides have similar absorbances. The similar UV spectrum obtained by photodiode arrays (PDA) detector in LC-MS analysis of TCP-sugars strengthen this idea (Fig. 4.11).

In figure 4.12 the HPLC spectrum of each TCP-glycosylation reactions using UDP-Glc or UDP-Ara as sugar donors are presented. As mentioned previously, TCP elutes at Rt: 15 min. Under the conditions tested 14 out of the 16 oat recombinant UGTs were able to form TCP-*O*-Glc (Rt 10.0 min). Only AsGT3i21 and AsGT11i11 did not catalyse the formation of this product (Fig. 4.12.A). The conversion of the TCP substrate into TCP-Glc is nearly complete for SAD10, AsGT16f23, AsGT24i2, AsGT23586A and AsGT23586B under the conditions used (Fig. 4.12.A). AsGT02436 also converts more than 50% of TCP (Fig. 4.12.A, top red spectrum). Under the conditions used AsGT14h20, AsGT14h21, AsGT25n16, AsGT27a12, AsGT27f7 and AsGT05827 produced TCP-Glc to a smaller extent, from 2 to 25% conversion; Fig. 4.12.A, table 4.3). Only traces of TCP-Glc were detected for AsGT21p16 and AsGT16h6 (Fig. 4.12.A, bottom blue and grey spectrums).
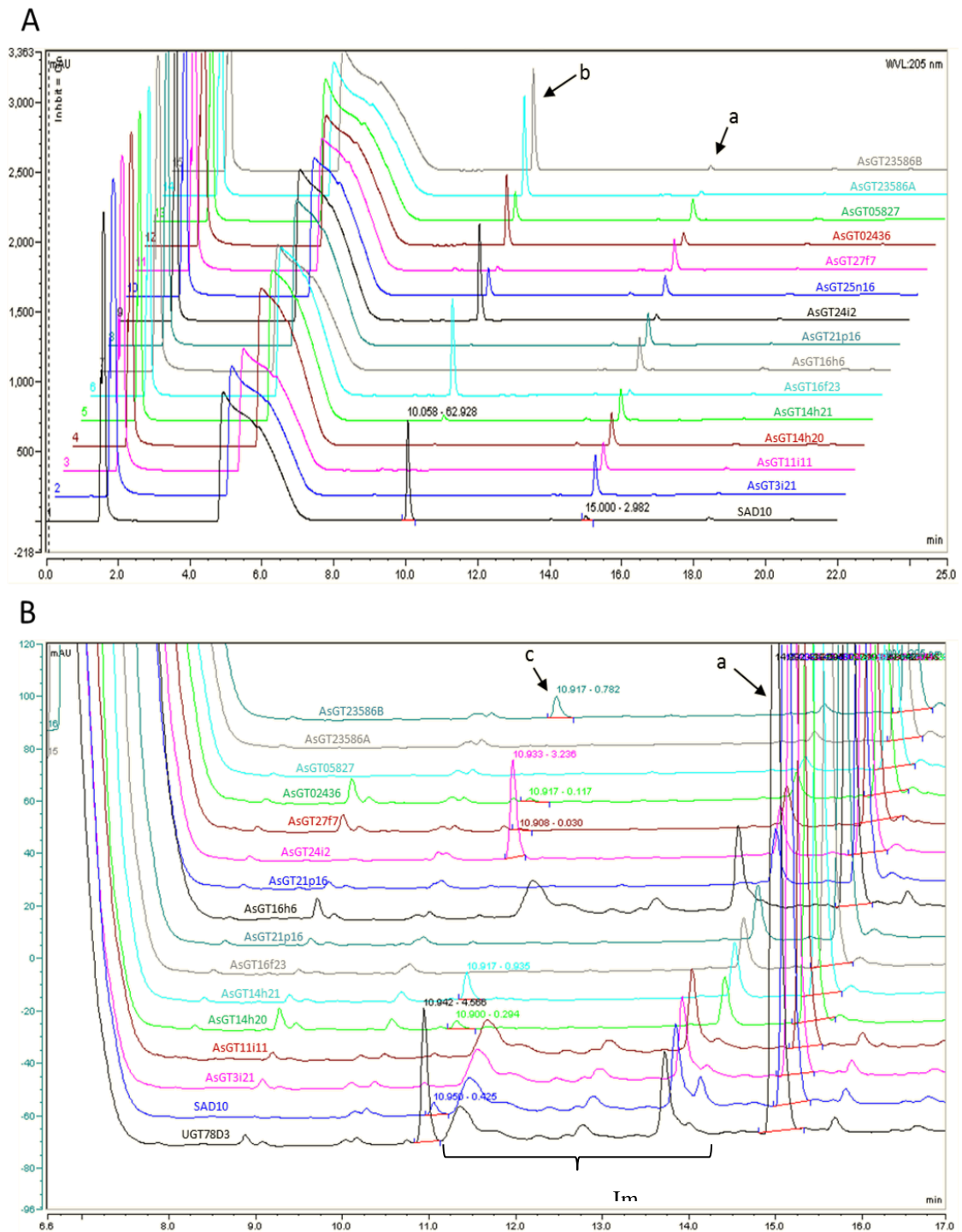
Figure 4.12: Comparison of TCP glycosylation following incubation with UDP-Glc (A) or UDP-Ara (B). Accumulation of TCP-glycosides was monitored by HPLC analyses at 205 nm after incubation for 1h at 30°C with 1 mM of UDP-sugar and 0.5 mM of TCP. Labelled peaks correspond to: **a**, TCP Rt 15.0 min; **b**, TCP-Glc Rt 10.1 min; **c**, TCP-Ara Rt 10.9 min; contaminant peaks are observed between peaks **a** and **c**.

Arabinosylation of TCP was catalysed by SAD10, AsGT14h20, AsGT14h21, AsGT24i2, AsGT25n16, AsGT02436 and AsGT23586B (Fig. 4.12.B). Most of these enzymes produce only traces of TCP-Ara - less than 1% conversion, except for AsGT24i2, which converted approximately 5% of TCP (Fig. 4.12.B, top pink spectrum). This can be compared with the positive control; UGT78D3 converts 7% of the TCP under these conditions (Fig. 4.12.A, bottom black spectrum).

Galactosylation of TCP was catalysed by SAD10, AsGT16f23, AsGT24i2, AsGT25n16, AsGT02436, AsGT05827, AsGT23586A and AsGT23586B (Fig. 4.13). Multiple peak products were detected in the galactosylation assays with the following retention times: Rt_1: 9.75min, Rt_2: 9.90 and Rt_3: 10.05 (labelled as **d1**, **d2** and **d3** in figure 4.13). Depending on the recombinant enzyme used, the predominant peak varied (Fig. 4.13); AsGT24i2 produces predominantly compound **d3** (Rt_3: 10.05min), and AsGT23586B produces compound **d2** as a major product (Rt_2: 10.05min); both of these enzymes are also producing the two other products in smaller quantities. The purity of substrates used has been assessed by proton NMR (UDP-Gal) and LC-MS (TCP) analysis (data not shown). No impurities or degradation products were detected. Additionally, proton NMR of UDP-Gal incubated with the enzyme preparation suggests that these odd product peaks are not due to sugar nucleotide epimerase activity present in the enzyme mix (data not shown). TCP degradation or modification in reaction conditions is unlikely due to absence of such multiple products with other UDP-sugars. Similar retention times suggest that the products have similar polarities, which conflicts with the notion that other sugar units have been added onto the acceptor. Formation and identity of these products remain elusive, consequently comparison with glucosylation and arabinosylation experiments is difficult. In subsequent figures only the highest peak produced by each enzyme was taken in consideration. Follow-up LC-MS analysis should bring new insight determining if the reaction products are structural isomers. The highest conversion is observed for AsGT23586B (approx. 12% conversion; Fig. 4.13). Approximately 4% conversion was observed for SAD10, AsGT16f23, AsGT24i2 and 23586A and 2.5% conversion for AsGT02436. Traces of **d3** were observed for AsGT25n16 and AsGT05827.
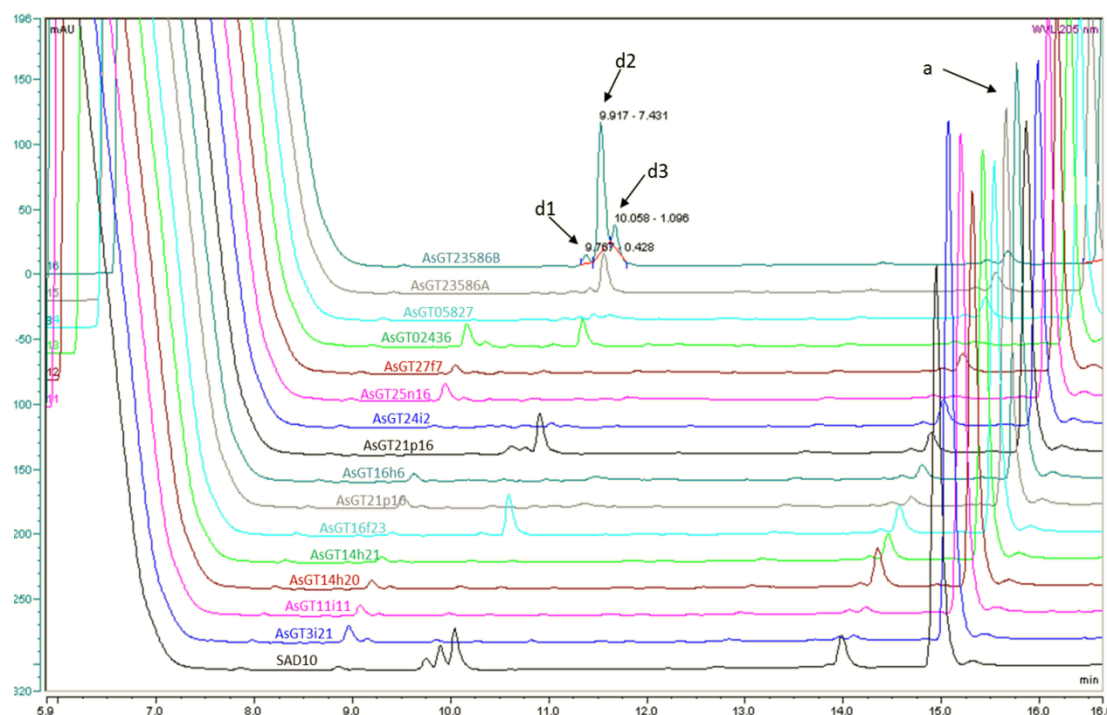
Figure 4.13: UGTs activity using TCP and UDP-Gal as substrates. SAD10 and AsGT16f23 (blue and grey spectrum) are active when incubated with TCP (Rt: 15 min) and UDP-Gal. SAD10 accumulates three products (Rt: 10.0, 9.9 and 9.8 min); AsGT16f23 accumulates a single product (Rt: 10.0).

The area under the product curve for the TCP-Ara and TCP-Gal reactions was compared to the area under the curve for the TCP-Glc reaction for each UGT. These results are summarised in table 4.3 and figure 4.14. None of the recombinant oat UGTs showed more activity with arabinose or galactose than they did with glucose. AsGT21p16, AsGT14h20 and AsGT14h21 had the highest arabinosylation activity towards TCP as compared to glucose, the area under the TCP-O-Ara curve representing 26%, 19% and 24% of that for TCP-O-Glc, respectively. None of the recombinant oat UGTs showed more than 5% galactosyltransferase activity compared to TCP glucosylation. The highest ratio of TCP-O-Gal to TCP-O-Glc formed was obtained with AsGT16f23.

Significantly, the TCP glycosylation experiments demonstrate that most of the recombinant UGTs have glucosyltransferase activity. Most of the UGTs tested also show a clear preference for UDP-Glc over UDP-Gal and UDP-Ara, using these latter UDP-sugars to only a limited extent, as observed in previous work on lamiales UGTs (Noguchi et al. 2009). Comparison between these two studies has to be made cautiously because Noguchi and co-workers studied sugar specificity onto two set of flavonoid UGT homologues in lamiales species, five flavonoid-7-*O*-glucosyltransferase homologues and five flavonoid-7-*O*-glucuronosyltransferase homologues. Their approach suggests that most of the glycosyltransferases tested used flavonoids as natural acceptors and consequently apigenin was used as substrate. In this study *A. strigosa* functions are unknown and TCP is potentially a poor substrate for the enzyme tested. Consequently, to push the system toward product formation and maximise our chance to detect glycosides, a 1:10 molar ratio of TCP to sugar donor was used. Under these conditions complete reaction is obviously reached for some enzymes (approx. 100% conversion with UDP-Glc for SAD10, AsGT16f23, AsGT24i2, AsGT23586A and AsGT23586B) introducing a bias in activity comparison between sugar donors (Fig. 4.14).
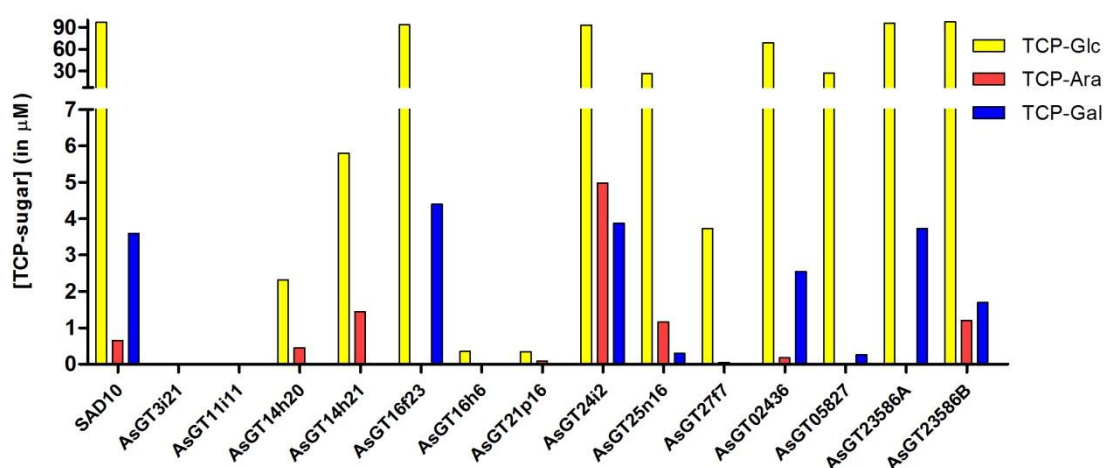
Figure 4.14: Comparison of TCP glycosylation catalysed by *A. strigosa* UGTs with three sugar donors. Yellow bars represent area under the curve of TCP-Glc; red bars represent area under the curve of TCP-Ara; blue bars represent area under the curve of TCP-Gal. Area under the curve is the integration of product peaks obtained at 205 nm.

Interestingly, some of the *A. strigosa* UGTs were able to use donors other than UDP-glucose. The arabinosylation activity of AsGT14h20 and AsGT14h21 towards TCP represents nearly a quarter of their glucosylation activity under the conditions used. Such a ratio is unusual when compared to other work (Kubo et al. 2004; Noguchi et al. 2009; Osmani et al. 2008) and might have significant implications for the catalytic capabilities of this enzyme *in vivo*. Activity testing with additional sugar donors, competition assays and kinetic analyses may help to improve our understanding of AsGT14h20 and AsGT14h21 sugar specificity.

Incubation of AsGT21p16 with UDP-Ara and UDP-Glc resulted in only traces of TCP-sugar formation. It is quite likely that the natural sugar donor has not been tested in the current study, or alternatively that TCP might be a very poor acceptor for AsGT21p16.

The inactivity of AsGT3i21 and AsGT11i11 under all conditions tested could be due to various factors: they may require sugar donors not tested in these experiments (e.g. UDP-xylose, UDP-rhamnose or UDP-glucuronic acid). AsGT3i21 and AsGT11i11 may be inactive due to misfolding or structural alterations occurring during the purification process; the Histidine tag may result in loss of activity; or they may be active but unable to use TCP as an acceptor substrate.

Overall, these results have to be considered cautiously because TCP is not the natural acceptor of the UGTs under investigation. Nevertheless, they provide insight into the abilities of the member of the *A. strigosa* UGT collection to utilise different sugar donors. From these experiments we can conclude that most of the UGTs tested are active glycosyltransferases and use UDP-Glc as their favoured sugar donor. Of note, arabinosylation activity was detected for six of the oat UGTs, in addition to the previously characterised SAD10 (Owatworakit et al. 2012). These enzymes are therefore candidates for triterpene arabinosylation.

Table 4.3: Summary table of recombinant UGTs activity toward TCP using various sugar donors.

UGTs	TCP-sugars formed (in µM)			Product ratio†	
	TCP-Glc	TCP-Ara	TCP-Gal	TCP-Ara/ TCP-Glc	TCP-Gal/ TCP-Glc
SAD10	96.8	0.65	3.60	0.7%	3.7%
AsGT3i21	-	-	-	n.d	n.d
AsGT11i11	-	-	-	n.d	n.d
AsGT14h20	2.32	0.45	-	20%	n.d
AsGT14h21	5.80	1.44	-	25%	n.d
AsGT16f23	93.4	-	4.40	n.d	4.7%
AsGT16h6	0.35	-	-	n.d	n.d
AsGT21p16	0.34	0.09	-	27%	n.d
AsGT24i2	93.3	4.98	3.88	5.4%	4.2%
AsGT25n16	26.3	1.16	0.30	4.4%	1.1%
AsGT27f7	3.73	0.05	-	1.2%	n.d
AsGT02436	68.6	0.18	2.55	0.3%	3.7%
AsGT05827	27.2	-	0.26	n.d	1.0%
AsGT23586A	95.8	-	3.73	n.d	3.9%
AsGT23586B	97.5	1.20	1.69	1.2%	1.7%
AsGT03999	-	-	-	n.d	n.d

* Area under the curve obtained at 205 nm was converted in concentration of products (in µM).

† The product ratio is the percentage of product formed with UDP-Ara or UDP-Gal compared to UDP-Glc

n.d value not determined

- No product detected

4.2.4   Glycosylation of triterpenes

4.2.4.1   Assays for triterpene glucosylation

Glucosylation assays were performed with all of the recombinant *A. strigosa* UGTs detailed above. UGT73C10, a *B. vulgaris* UGT with regiospecific 3-*O*-glucosylation activity onto oleanane-type triterpenoid acceptors, was used as a positive control for the formation of triterpenoid glucosides. Glucosylation assays were performed using β-amyrin, hederagenin and oleanolic acid as potential oleanane-type triterpene acceptors (Fig. 4.15). Due to the hydrophobicity of oleanane type triterpenes 10-20% of ethanol was systematically used as a co-solvent to solubilise those chemicals in following enzymatic assays (20% for β-amyrin, 10% for hederagenin and oleanolic acid).
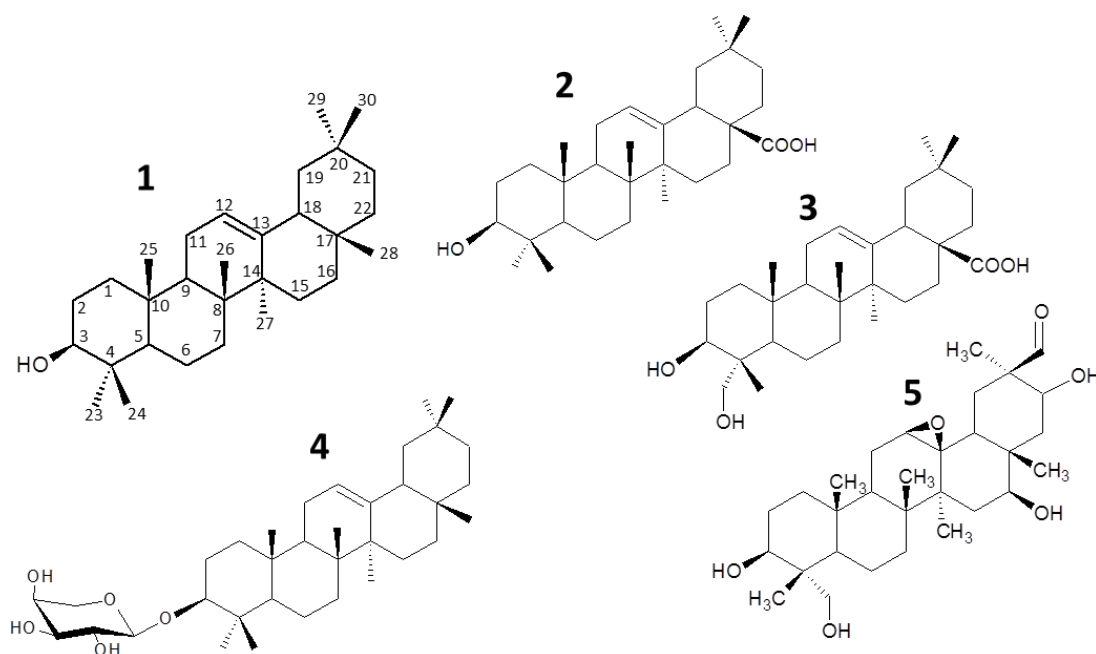
Figure 4.15: Structure of triterpenoids used as acceptors. β-Amyrin (1), oleanolic acid (2), hederagenin (3), β-Amyrin-3-*O*-α-L-arabinose (4), The expected natural substrate of avenacin glycosyltransferases is shown for information (5).

Under the conditions used for these assays, the positive control, UGT73C10 catalysed the formation of products with each of the oleanane-type triterpenes tested, as already reported by Augustin et al. (2012) (method in section 2.2.14). According to Augustin et al. (2012) these products are 3-*O*-glucosides of their relative acceptors. Under identical conditions, thin layer chromatography analysis led to detection of a new compound with recombinant AsGT02436 in the presence of hederagenin (Fig. 4.16.A). Similar separation profiles suggest that the AsGT02436 product might be hederagenin monoglucoside. Three positions on hederagenin could in principle be glucosylated by an *O*-glucosyltransferase, namely C-3, C-28 or C-23 (Figure 4.16.B).Glycosylation at the C-3 position is the most common for saponins (Vincken et al. 2007).

This compound could not be detected by mass spectrometry in conditions used in JIC metabolic services (see section 2.2.25); therefore, the formation of monoglucosylated hederagenin could not be confirmed. Simple sterols and related compounds are known to ionised poorly, therefore those compounds are generally derivatised prior GC-MS analysis (Geisler et al. 2013; Lagarda et al. 2006). In the case of steryl glycosides, analysis usually relies on acidic hydrolysis then derivatisation of the aglycone (Nystrom et al. 2012). Alternatively silylation of the intact compound can be performed for GC-MS analysis (Phillips K.M et al. 2004; White et al. 2010). In the present study, hydrolysis of the product to remove the glycosidic moiety before analysis was not suited for the confirmation of glycosylation activity; consequently, this approach was not intended. Silylation of the reaction product results in detection of the aglycone alone after potential sugar loss imputable to the process. Alternatively, assessment of glucose ester formation (glucosylation on C-28 position) may be investigated with basic hydrolysis; this treatment would leave glycosidic bounds intact but hydrolysed ester bounds. No other triterpene glucosyltransferase activity was detected under the conditions used for any of the other *A. strigosa* UGTs (*in vitro* assays results summary in table 4.4).
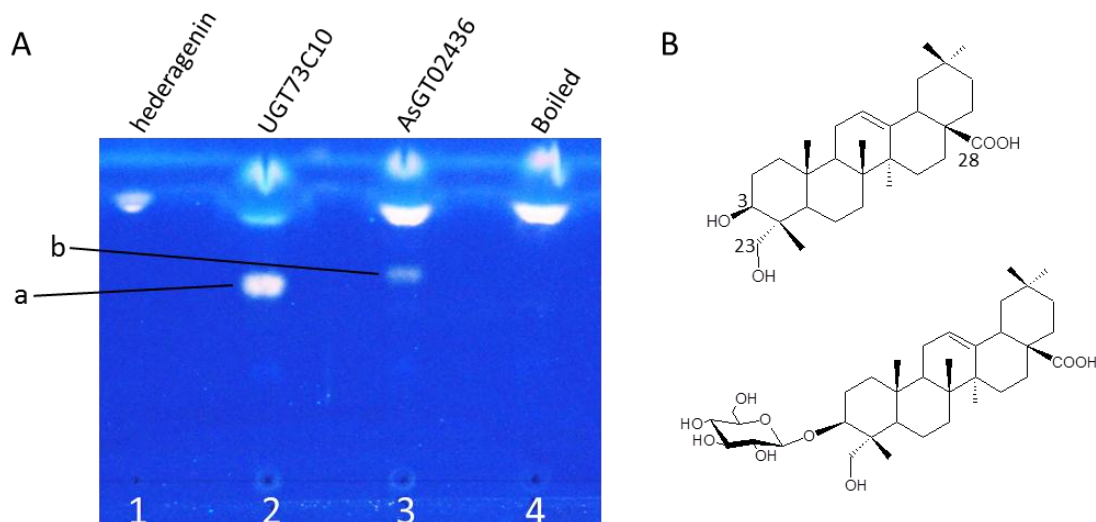
Figure 4.16: Comparison of glucosyltransferase activities of UGT73C10 and AsGT02436 towards hederagenin. A) Thin layer chromatography of hederagenin glucosylation assays with purified recombinant UGTs. Reaction products were extracted with ethyl acetate. The TLC was stained with methanol:sulfuric acid (9:1), after heating the image was taken under UV illumination. Lane 1, hederagenin standard; lane 2, generation of glucosylated hederagenin catalysed by UGT73C10; lane 3, generation of a UDP-Glc + hederagenin reaction product by AsGT02436; lane 4, reaction control with boiled AsGT02436. The arrows "a" and "b" show the products of each reaction. **B.** Chemical structures of hederagenin (top) and hederagenin-3-*O*-glucose (bottom).

Chemical synthesis of β-amyrin-3-*O*-α-L-arabinose (β-AA) was carried out by the Field laboratory at JIC. This compound was used as an acceptor for glucosylation assays in order to identify avenacin glucosyltransferases that are able to add β-1→4 and β-1→6 linked D-glucose molecules onto the L-arabinose residue. In the scope of identifying UGTs involved in avenacin biosynthesis. TLC analysis suggests that none of the *A. strigosa* UGTs recognised *β*-AA as a substrate. Only *β*-AA was detected on stained TLCs; an example of TLC analysis is presented in figure 4.17.
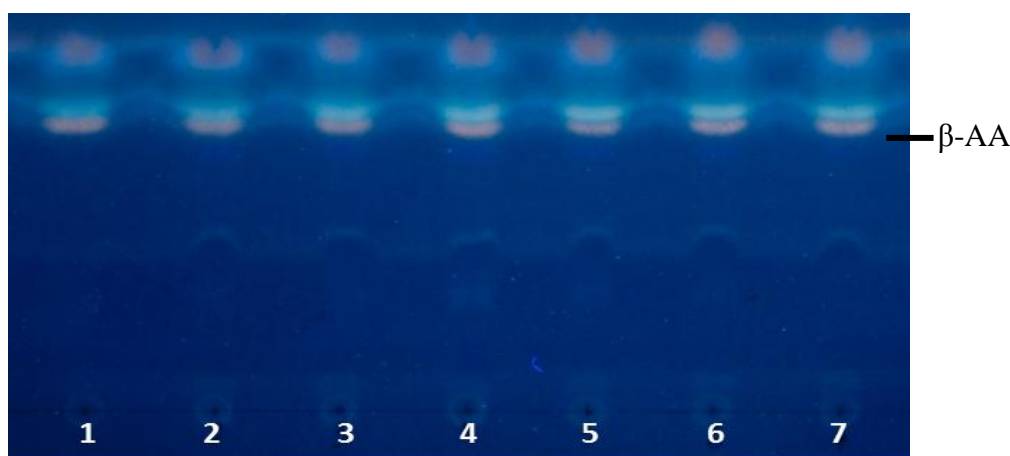
Figure 4.17: Glucosylation assay of β-amyrin-Ara. Thin layer chromatography of β-amyrin-3-*O*-*β*-L-arabinose glucosylation assays with purified recombinant UGTs. Reaction products were extracted with ethyl acetate. The TLC was stained with methanol:sulfuric acid (9:1) and compounds visualised under UV illumination. Substrates were incubated without enzyme (lane 1), AsGT24i2 (lane 2), AsGT05827 (lane 3), AsGT23586A (lane 4), AsGT23586B (lane 5), AsGT14h20 (lane 6) and AsGT16f23 (lane 7).

4.2.4.2   Assays for arabinosylation activity towards triterpenoids

Arabinosylation assays were performed with SAD10, AsGT14h20, AsGT14h21, AsGT24i2, AsGT25n16, AsGT02436 and AsGT23586B, since these *A. strigosa* UGTs had previously been shown to catalyse formation of TCP-Ara (table 4.3; method in 2.2.14). AsGT3i21 and AsGT11i11 were also included in this experiment considering the absence of information regarding their sugar specificity (e.g. absence of activity toward TCP). Enzymatic assays were performed using similar conditions to those used for glucosylation assays (Fig. 4.18). Various conditions were tested for these enzymes, but from the entire UGT collection (temperature, substrate concentration, solvent to help solubilised the acceptor) none of these conditions led to detection of a product.

The limited number of acceptors available and the difficulties encountered with mass spectrometry analysis of the reaction products were limiting factors for the investigation of triterpenoid glycosyltransferase activity. The systematic use of 10-20% of ethanol in our assays to solubilise β-amyrin or hederagenin may also be detrimental for the activity of some of our UGTs despite the activities detected in control reactions (UGT73C10) and AsGT02436. Enzymatic assays with dimethyl sulfoxide or cyclodextrin did not show more activity; glycosyltransferase enzymatic assays have been reported with both co-solvents (Moraga et al. 2004; Nagatoshi et al. 2012; Wang et al. 2013b).
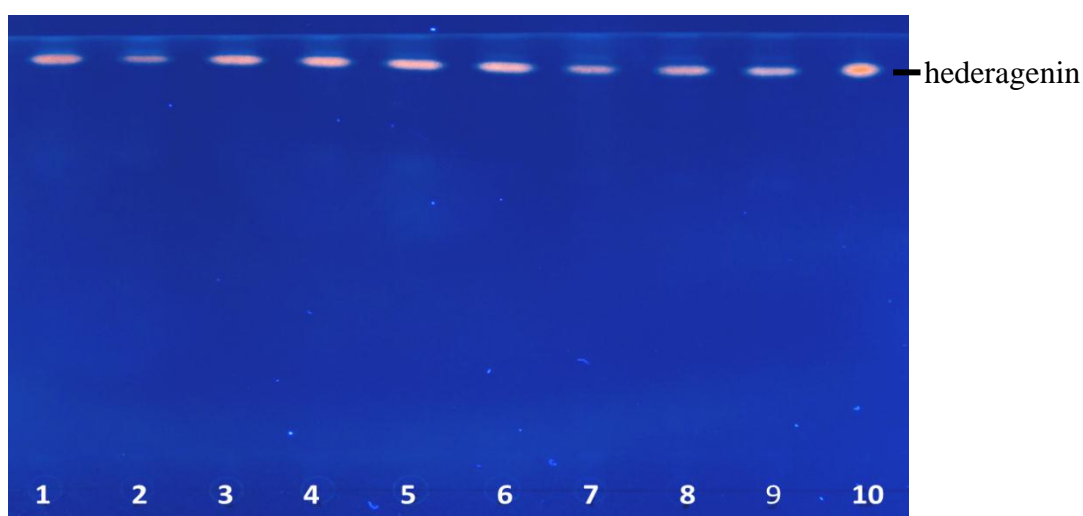
Figure 4.18: Arabinosylation assay of hederagenin. Thin layer chromatography of hederagenin arabinosylation assays with purified recombinant UGTs. Reaction products were extracted with ethyl acetate. The TLC was stained with methanol:sulfuric acid (9:1) and compounds visualised under UV illumination. Reactions products are loaded from lane 1 to 9, hederagenin standard is in lane 10. Reactions were performed with AsGT3i21 (1), AsGT11i11 (2), AsGT14h20 (3), AsGT14h21 (4), AsGT21p16 (5), AsGT24i2 (6), AsGT25n16 (7), AsGT23586B (8), UGT73C10 (9).

acceptors \ sugar donors	UDP-Glc	UDP-Ara
β-amyrin	no activity	no activity
β-amyrin-3-*O*-arabinose	no activity	no activity
hederagenin	AsGT02436	no activity
Oleanolic acid	AsGT02436	no activity

Table 4.4: Table summarising the results obtained from in vitro assay of *A. strigosa* UGTs over triterpenoids. All 14 recombinant UGTs have been used in glucosylation assays (see table 4.1). Enzymes capable to utilised UDP-Ara in TCP assays (see section 4.2.3) were used in arbinosylation assays (e.g. AsGT14h20, AsGT14h21, AsGT21p16, AsGT24i2, AsGT25n16, AsGT23586B), AsGT3i21 and AsGT11i11 have been added to those test considering the absence of activity observed with TCP and their distant phylogenetic relationship to group D (Fig. 3.1). All glucosylation assays have been tested with UGT73C10 and were all positives except for β-amyrin arabinoside as expected.

The absence of candidates showing a preference for UDP-Ara over UDP-Glc in the TCP experiment (see section 4.2.3.3), combined with the lack of arabinosyltransferase activity towards triterpene acceptors, suggests that avenacin arabinosyltransferases have not been cloned in this study. However it is possible that the substrates used were not appropriate for the identification of avenacin glycosyltransferases. Accumulation of β-amyrin in *sad2* mutants suggests the addition of L-arabinose to the triterpene scaffold during avenacin biosynthesis occurs after modification of β-amyrin by the CYP450 SAD2 (Qi et al. 2004). Additional oxidations (C-21, C-23 and C30) might take place before the sequential addition of the sugar units to form the avenacin trisaccharide (hypothetical avenacin pathway, Fig. 1.11). Accumulation of desacyl avenacin in *sad7* mutant suggests completion of trisaccharide formation is a prerequisite to the acylation catalysed by SAD7 (Mugford et al. 2009). Consequently, modified β-amyrin backbone may be required for recognition of the triterpene acceptor by triterpene-modifying UGTs.

4.3 Conclusion

In this chapter, I report the cloning of 14 oat UGTs into expression constructs for expression in *E. coli* and *N. benthamiana* leaves. Three functionally characterised enzymes were included as positive controls. These were the oat UGT SAD10, which had previously been shown to glucosylate *N*-methyl anthranilate, the acyl donor used by the acyltransferase SAD7 (Owatworakit et al 2013); the flavonol-3-*O*-arabinosyltransferase UGT78D3 from *A. thaliana* (Yonekura-Sakakibara et al. 2008); and UGT73C10 from *B. vulgaris,* which has previously been shown to glucosylate oleanane-type triterpenes (Augustin et al. 2012). Expression and purification of recombinant oat UGTs resulted in active enzymes in most cases, as demonstrated by the activities detected towards the generic substrate TCP. Soluble expression of AsGT03999 was not achieved. AsGT03999 is one of the UGT80 plant sterol glycosyltransferases, which are known to have an *N*-terminal anchor to the membrane generally responsible for insolubility of recombinant proteins. Some reports have mentioned the solubility and activation of sterol glycosyltransferases by non-ionic detergents; this approach has not been attempted here (Grille et al. 2010; Warnecke et al. 1997). Only small quantities of soluble enzymes were obtained for AsGT3i21 and AsGT11i11. Those enzymes present some similarities (61% identity) and did not belong to any phylogenetic group characterised so far (Fig. 3.1). The apparent insolubility of these enzymes may explain the complete absence of activity observed towards acceptors tested using soluble *E. coli* protein fractions enriched with ion immobilized affinity chromatography.

4.3.1 Glucosyltransferase activity is predominant in *A. strigosa* UGT collection

A method was designed to determine sugar specificity for UGTs of unknown function using TCP as an acceptor. This method had the advantage that it did not rely on radiolabelled sugar donors. Assays with the glucosyltransferase SAD10 and the arabinosyltransferase UGT78D3 validated the method by showing that sugar specificity was retained with TCP when compared to the reactions for the natural acceptors of these two enzymes. This method was then applied to the oat UGT collection. All of the *A. strigosa* UGTs able to use TCP as an acceptor showed a preference for UDP-Glc over UDP-Ara and UDP-Gal. This is in accordance with the vast majority of plant UGTs (Bowles et al. 2006; Osmani et al. 2009; Wang 2009). This is illustrated by the subset of functionally characterised plant glycosyltransferases included in the phylogenetic analysis in section 3.2.2, containing 85% of glucosyltransferases (Fig. 3.1). None of the *A. strigosa* UGTs shows a preference for UDP-Ara over UDP-Glc; however 6 of these enzymes were able to use UDP-Ara as a substrate and were tested as potential triterpenoid arabinosyltransferases. The majority of recombinant *A. strigosa* UGTs tested catalyse glucoconjugation of TCP, supporting the idea that UGTs are able to catalysed glycosylation of both endogenous and xenobiotic substrates. Messner et al. (2003) performed competition studies using both natural substrate and TCP over recombinant UGTs and demonstrated that glycosyltransferase activity over endogenous native substrates of some enzyme was inhibited in the presence of the xenobiotic TCP. Messner and co-workers suggested that xenobiotics are not conjugated by specialised detoxifying glycosyltransferases but processed by UGTs active on endogenous substrates expressed in the contaminated tissue.

4.3.2   Implication of *in vitro* assays for avenacin biosynthesis

Attempts to validate *in vitro* assays with triterpene acceptors (e.g. β-amyrin, hederagenin, oleanolic acid) using UGT73C10, a hederagenin-3-*O*-glucosyltransferase from *B. vulgaris* gave results in accordance with those of Augustin et al. (2012), confirming that the conditions used were suitable for assaying triterpene glycosylation. Recombinant AsGT02436 formed a product when hederagenin was used as a substrate. The retention time of this compound suggests that the AsGT02436 product is likely to be a hederagenin monoglucoside. Scale up and NMR analysis is required to formally identify this compound and hence to shed light on the regioselectivity of AsGT02436 towards triterpenoids.

Arabinosylation is the proposed first step toward the sequential synthesis of avenacin sugar moiety. Arabinosyltransferase activity was not detected using oleanane type triterpenoids as acceptors (table 4.4). Nevertheless, the structure and physical properties of acceptors used are divergent from potential natural acceptor (e.g. deglycosylated desacylavenacin). Production of optimal substrate requires chemical or enzymatic approaches to deglycosylated avenacin (Augustin et al. 2012; Osbourn et al. 1995; Shibuya et al. 2010) or desacyl avenacin accumulated in *sad7* mutant lines (Mugford et al. 2009). Synthesis of such a compound will also help solubilisation issues encountered with triterpene acceptors. The absence of enzyme showing preference for UDP-Ara in TCP assays together with the inability to form triterpene arabinoside with β-amyrin or hederagenin, might suggests avenacin arabinosyltransferase has not been cloned in the present study. Avenacin glycosyltransferases may be part of the few candidate genes remaining to be amplified (see section 4.2.1). Alternatively, the strategy developed in chapter 3 might have been inefficient to uncover avenacin glycosyltransferase candidates. The absence of reference genome to improve the quality of 454 reads assembly, or the lack of transcriptomic data from avenacin non-synthesising oat tissue allowing transcript abundance comparison are detrimental for the optimal use of the 454 generated data. Additionally, extension of the oat genome sequencing around the avenacin core cluster may help delineate essential UGT candidates.

The absence of enzyme favouring UDP-Ara over UDP-Glc as sugar donor is preoccupying. As mentioned above the hypothetical model of avenacin trissacharide formation consist of sequential addition of the sugar units catalysed by members of the UGT family. Such a scenario has been observed for the assembly of all of the saponin glycosidic moieties reported in literature (Achnine et al. 2005; Augustin et al. 2012; Itkin et al. 2013; Kohara et al. 2005; Kohara et al. 2007; Meesapyodsuk et al. 2007; Naoumkina et al. 2010; Sayama et al. 2012; Shibuya et al. 2010). Alternative glycosylation mechanism may occur in avenacin biosynthesis involving other enzyme families. A recent report mentioned a transglucosidase from the glycoside hydrolase family one (GH1) catalysing glucose transfer from phenolic acid esters to various natural compound and targeted to vacuoles in rice (Luang et al. 2013). Authors suggest this enzyme may be involved in homeostasis of plant metabolites. Blockwise addition of a preformed activated trisaccharide may be an alternative route toward avenacin biosynthesis; such a hypothesis may be address by investigation of catalytic properties of oat protein extract over suitable substrates.

4.3.3 General conclusion

In this chapter, new approaches were used to decipher plant glycosyltransferase activities. The generic substrate of plant UGTs, TCP, was used as an acceptor to determine sugar donor specificities of glycosyltransferases for which natural acceptors are unknown. Comparative analysis of activity with three sugar donors was conducted to determine the sugar donor preference of each recombinant UGT. *In vitro* assays with recombinant *A. strigosa* UGTs led to the identification of hederagenin glucosyltransferase activity catalysed by AsGT02436. Kinetic analysis of this enzyme over hederagenin and other potential acceptors has to be done to understand substrate preference and possible biological function.

Chapter 5 - A combinatorial approach to synthesised new-to-nature triterpenoids in *Nicotiana benthamiana*

5.1 Introduction

The strategy developed in the previous chapter for the characterisation of triterpenoid glycosyltransferase activity presents some limitations. The limited number of oleanane-type triterpenoids readily available and the complexity of accessing the avenacin aglycone (e.g. epoxide group unstable in acidic conditions) make *in vitro* approach delicate. The low solubility of triterpenoid acceptors and the promiscuity of plant UGTs are also limiting factors when it comes to evaluation of the biological significance of experimental data. The development of the pEAQ vector system allows transient co-expression of multiple proteins in *Nicotiana benthamiana* leaf tissue. Osbourn and co-workers have shown the value of this system for heterologous synthesis of secondary metabolites (Geisler et al. 2013; Mugford et al. 2013).

5.1.1 Origin and development of the pEAQ vector series

The development of the pEAQ vector system by the Lomonossoff laboratory at the John Innes Centre has opened up new opportunities for rapid tests of function of uncharacterised enzymes in *N. benthamiana* using transient expression methods (Sainsbury et al. 2009). The pEAQ vectors are based on cowpea mosaic virus (CPMV), a bipartite RNA virus comprising RNA-1, which encodes the viral replication machinery, and RNA-2 which encodes the viral coat proteins and movement proteins. The expression system developed by Lomonossoff and co-workers does not rely on viral replication but uses a deleted version of RNA-2 to achieve high expression yield of heterologous proteins. The gene of interest is inserted between the 5' and 3' UTRs of the CPMV RNA-2. The modified 5' UTR sequence and the co-expression of the suppressor of RNA silencing P19 provide high stability and enhance translation of the sequence flanked by UTRs, allowing high expression level without RNA-1 (Sainsbury and Lomonossoff 2008). The insertion of P19 and CPMV UTR cassette into Agrobacterium T-DNA leads to the

development of pEAQ vectors. These vectors allow the transient expression of the desired protein in *N. benthamiana* agroinfiltrated tissues. The advantages of the pEAQ system are that a high protein yield is reached in only a few days, combined with the simplicity of the Agrobacterium-mediated transformation system.

5.1.2   The pEAQ vector series: A convenient system for heterologous expression of proteins

The pEAQ system has been used successfully for a variety of purposes. The production of virus-like particles has been particularly well developed due to applications in immunology or for other biotechnological purposes (Peyret and Lomonossoff 2013). The pEAQ vectors have also been used to produce active enzymes; human gastric lipase was produce with a high expression level of 0.5 g/kg fresh weight of *N. benthamiana* leaves. Functional analysis of chitinase from rice and sesquiterpene synthases from *Artemisia annua* has also been achieved using this expression system (Kanagarajan et al. 2012; Miyamoto et al. 2012). Yields of ca 100 mg/kg were obtained for the heterologous expression of the two sesquiterpene synthases.

The Osbourn lab has previously shown the utility of the pEAQ system to reconstitute metabolic pathways in a heterologous plant system. Transient co-expression of SAD9 and SAD10 led to the accumulation of the acyl glucose donor *N*-methylanthraniloyl glucose (Mugford et al. 2013). SAD7, the acyltransferase that uses this activated acyl donor to acylate the triterpene scaffold during avenacin synthesis has also been expressed in functional form in *N. benthamiana* (Mugford et al. 2009). Co-expression of oat β-amyrin synthase (SAD1) and the CYP450 that modifies the β-amryin scaffold (SAD2) using the pEAQ expression system resulted in the accumulation of a new product that could be purified from *N. benthamiana* leaves in milligram quantities, enabling the structure of this modified product to be determined by NMR spectroscopy (Geisler et al. 2013).

5.1.3 Aims

In this chapter, heterologous co-expression of UGTs with SAD1 and SAD2 in *N. benthamiana* using the pEAQ system was achieved. Extraction and purification of the first glycosylated triterpenoid produced in *N. benthamiana* may be regarded as a proof of principle of the relevance of this system for the production of new-to-nature triterpenoids. The present system offers us a powerful tool for the discovery of uncharacterised triterpenoid glycosyltransferases.
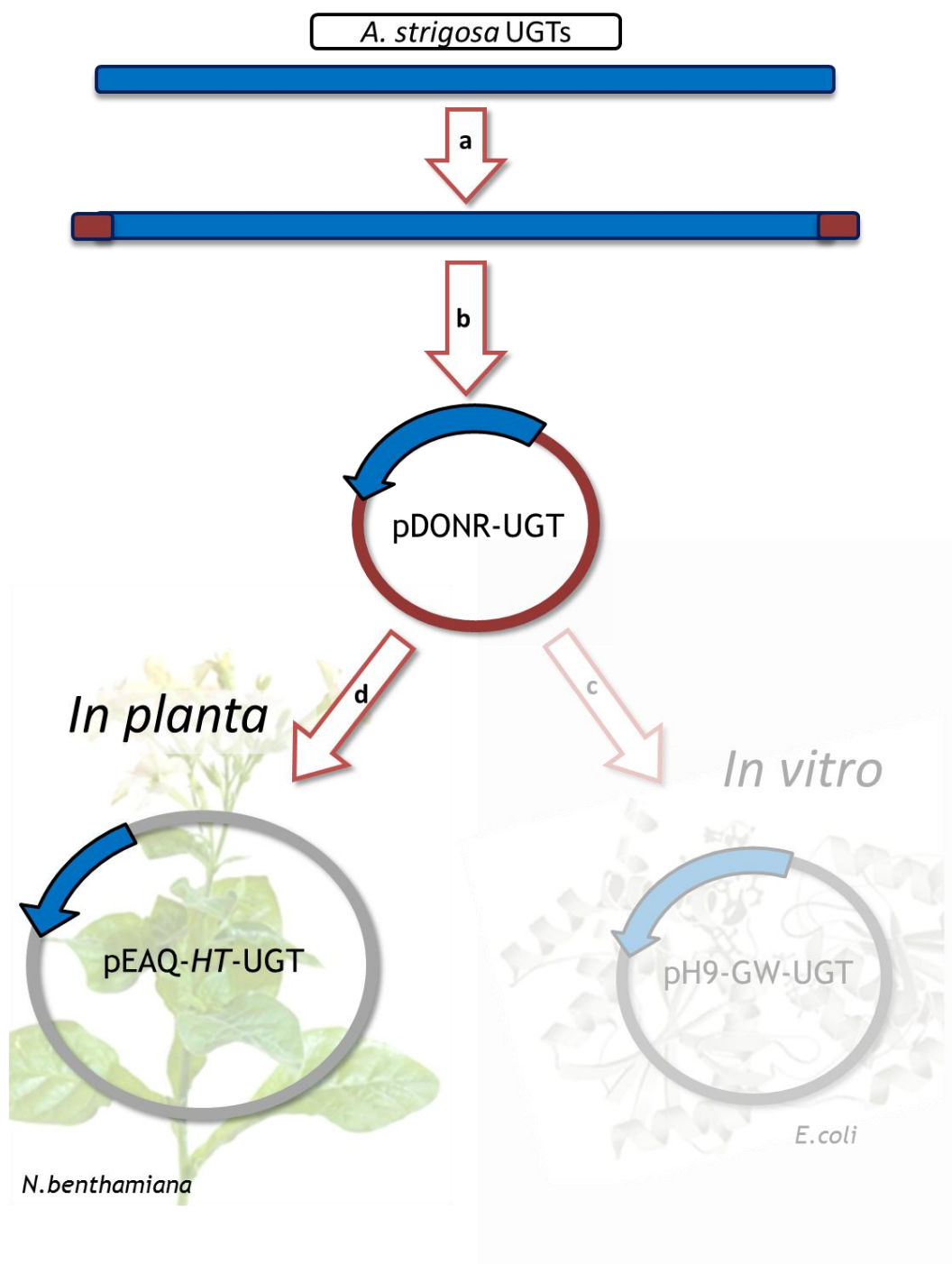
Figure 5.1: Overview of the strategy for cloning *A. strigosa* UGT coding sequences into PEAQ vector for transient expression in *N. benthamiana*. A two-step PCR protocol was used to amplify the UGT coding sequences and add terminal AttBs adapters (a). The UGT coding sequences were then inserted into the GATEWAY entry vector pDONR207 using the BP clonase reaction (b) (Hartley et al, 2000). Two expression vectors were constructed from each pDONR construct using LR clonase reactions (Hartley et al, 2000). The expression vector pEAQ-*HT*-Dest1 was used for expression of UGTs in *N. benthamiana* leaf tissues by *Agrobacterium*–mediated transient expression (Sainsbury et al, 2009) (d).

5.2 Results and discussion

5.2.1 Expression of SAD1 and SAD2 in leaves of *N. benthamiana* using pEAQ system

In previous work to elucidate SAD2 activity, SAD1 was either expressed alone or with SAD2 in transient expression experiments in *N. benthamiana* leaves (Geisler et al. 2013). The constructs used are shown in figure 5.2.A. GC-MS analysis confirmed the accumulation of β-amyrin in SAD1-expressing tissues (Fig. 5.2.B). When SAD2 was co-expressed with SAD1, β-amyrin level decreases in agroinfiltrated tissues and a new compound was produced (Fig. 5.2.B). This compound was identified as 12,13β-epoxy-16-β-hydroxy-β-amyrin, indicating that SAD2 is able to perform C16 hydroxylation and C12,13 epoxidation of β-amyrin. These experiments demonstrated that endogenous 2,3-oxydosqualene can be efficiently converted into β-amyrin in *N. benthamiana* leaves, so providing a basis for bioengineering of triterpenoid pathways using this system. Co-expression of SAD1 and SAD2 led to accumulation of approximately 1.2 mg/g of SAD2 product from dry leaf material. Since this system enables the early triterpene pathway intermediates to be generated by co-expression of pathway enzymes it was used, in this study, as a basis for developing an *in planta* platform for functional characterisation of *A. strigosa* UGTs.

Accumulation of SAD1 and SAD2 proteins in *N. benthamiana* leaves following agroinfiltration is optimal 6 days after infiltration (Geisler et al. 2013). In the present study, Western blot analysis showed that SAD10 was also expressed after 6 days (Fig. 5.3; methods in section 2.2.19) through detection using a specific anti-SAD10 antibody (Owatoworakit et al 2013). Co-expression of the three saponin biosynthetic enzymes (oxidosqualene cyclase (SAD1), P450 (SAD2) and glycosyltransferase (*A. strigosa* UGT collection) in the same tissues at the same time is a prerequisite to analyse modification of the triterpene backbone by oat UGTs.
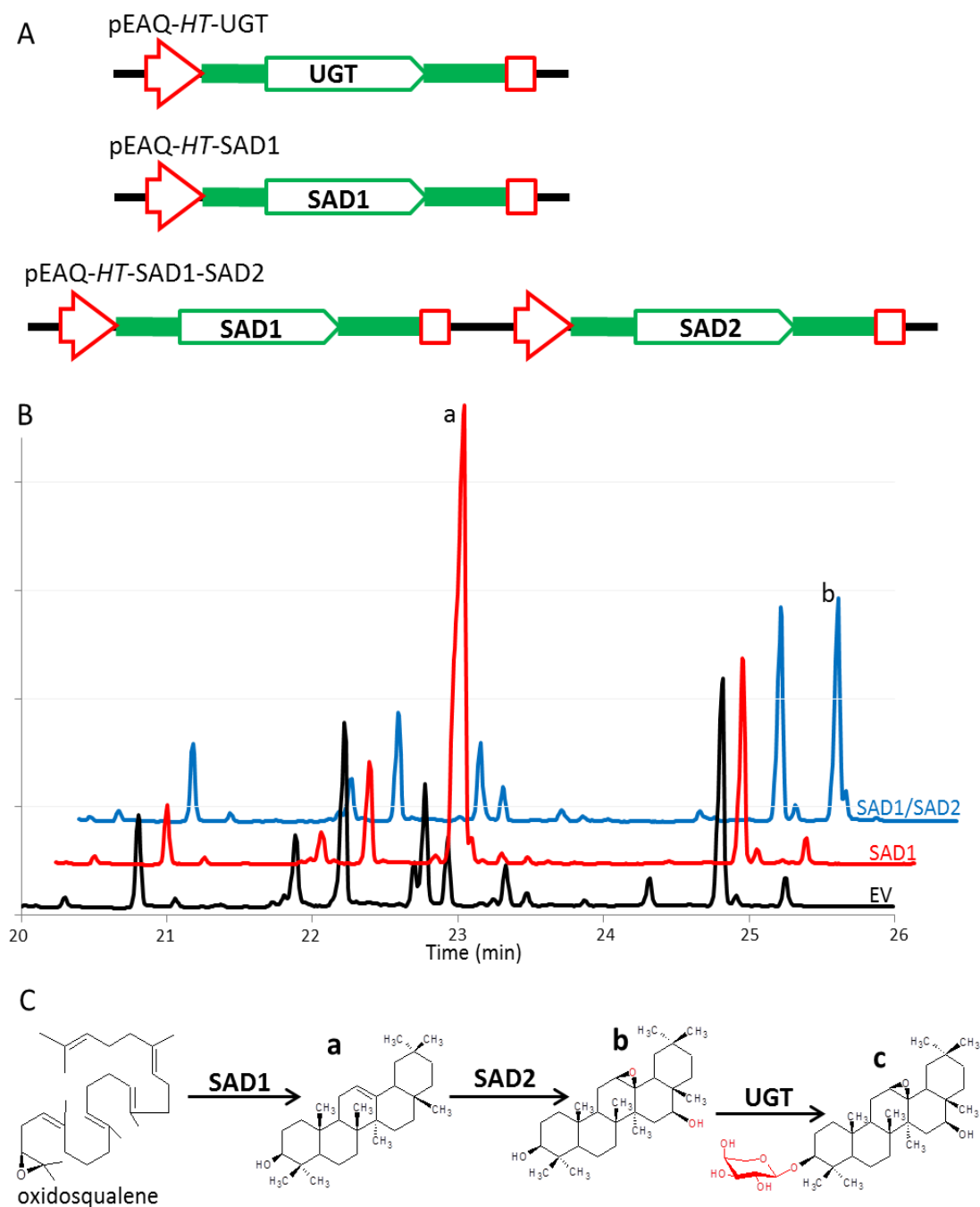
Figure 5.2: Co-expression of multiple enzymes in *N. benthamiana* for the creation of a saponin pathway. A. The pEAQ-*HT* based constructs for the transient expression of SAD1, SAD1+ SAD2 or *A. strigosa* UGTs. Red arrow, promoter 35S; red box, Nos terminator; thick green lines, CPMV RNA2 UTR; green arrow, CDS. pEAQ-*HT*-SAD1 and pEAQ-*HT*-SAD1-SAD2 were constructed previously (Geisler et al 2013) **B.** GC-MS analysis performed by K. Geisler showing the accumulation of triterpenoids in *N. benthamiana* tissues infiltrated with empty vector (black); pEAQ-*HT*-SAD1 (red) and pEAQ-*HT*-SAD1-SAD2 (blue). Accumulation of β-amyrin (a) is detected in tissues expressing SAD1. Accumulation of 12,13-epoxy-16-hydroxy-β-amyrin (b) is detected in tissues expressing SAD1 and SAD2. The identification of this compound is detailed in supplementary data S.11. **C.** Co-expression of triterpene glycosyltransferases with SAD1 and SAD2 is expected to lead to the accumulation of 12,13-epoxy-16-hydroxy-β-amyrin-3-*O*-α-L-arabinose (c).
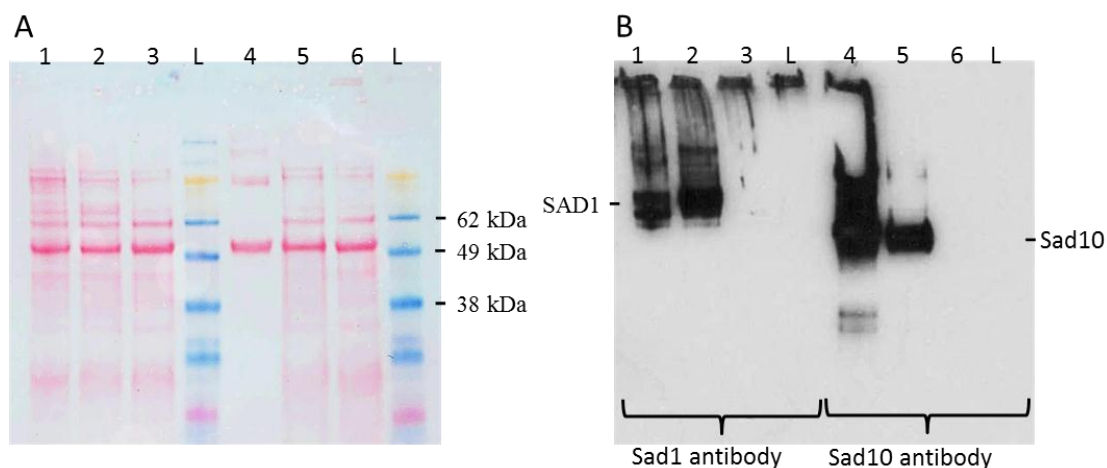
Figure 5.3: Comparative analysis of SAD1 and SAD10 expression in *N. benthamiana* leaf tissues 6 days after agroinfiltration. Western blot was performed with SAD1 antibody (Lanes 1-3) and SAD10 antibody (Lanes 4-6). SDS-PAGE was loaded with approximately 20µg of protein extracted from tissues infiltrated with the following constructs: lane 1, pEAQ-*HT*-SAD1; lane 2, pEAQ-*HT*-SAD1-SAD2; lane 3, non-infiltrated plant; lane 4, purified recombinant SAD10; lane 5, pEAQ-*HT*-SAD10; lane 6, non-infiltrated plant; lanes L, SeeBlue® Plus2 Pre-Stained Standard. **A.** Ponceau stained blotted membrane. **B.** Film exposed for 1 min reveals peroxidase activity from secondary antibody.

5.2.2 Co-expression of UGT73C10 with SAD1 and SAD2 leads to the accumulation of glycosylated triterpenoid

5.2.2.1 *Co-expression of the three proteins in N. benthamiana leaves*

In order to evaluate the feasibility of glycosylating SAD1 and SAD2 triterpenoid products, we co-expressed SAD1, SAD2 and the hederagenin 3-*O*-glucosyltransferase UGT73C10 in *N. benthamiana* leaves using the corresponding pEAQ vectors for agroinfiltration (e.g. pEAQ-*HT*-SAD1-SAD2/pEAQ-*HT*-SAD1 and pEAQ-*HT*-UGT73C10; see Fig. 5.1). In addition to hederagenin, UGT73C10 is able to catalyse 3-*O*-glucosylation of β-amyrin, oleanolic acid and betulinic acid *in vitro* (Augustin et al. 2012). Transcript analysis of non-infiltrated and agroinfiltrated *N. benthamiana* leaves revealed the co-expression of UGT73C10 with SAD1 and SAD2 (Fig. 5.4) in the tissues infiltrated with pEAQ-*HT*-SAD1-SAD2 and pEAQ-*HT*-UGT73C10. Kanamycin and UTR specific primers were used to control the presence of DNA contamination (T-DNA) in the experiment (see section 2.2.7).
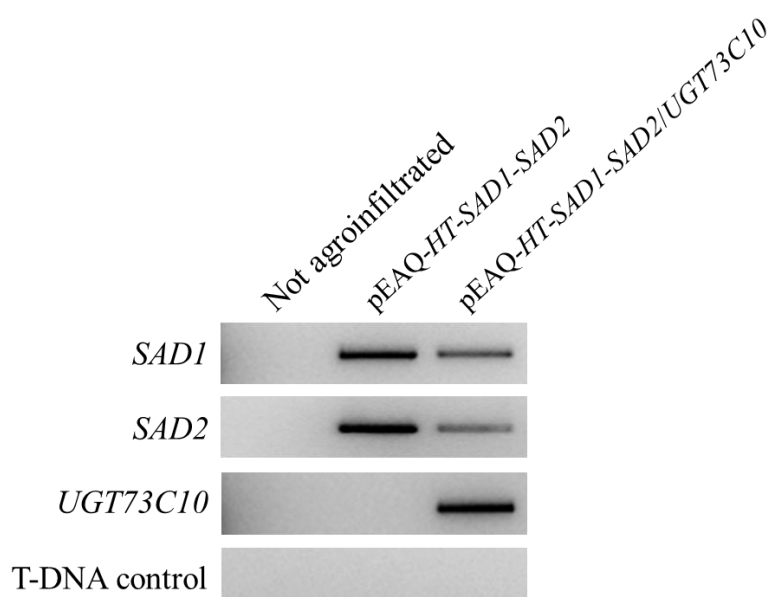
Figure 5.4: Transcript accumulation of UGT73C10, SAD1 and SAD2 in *N. benthamiana* agroinfiltrated tissues. Expression analysis was conducted using mRNA-reverse transcription-PCR (RT-PCR) technique. Total RNA was extracted from *N. benthamiana* agroinfiltrated tissues, 6-days after infiltration. A cDNA was synthesised using primer specific for RNA-2 5' untransformed region (UTR, specific of the pEAQ transcripts, see Sainsbury et al. 2008).

5.2.2.2 *UGT73C10 expression leads to the accumulation of new compounds*

TLC analysis of methanolic extracts of agroinfiltrated tissues is consistent with the results of Geisler et al. (2013) (Fig. 5.5). The accumulation of β-amyrin (**a**) is observed in tissues expressing SAD1 compare to uninfiltrated control (lanes 1-3, Fig. 5.5.A). In tissues co-expressing SAD1 and SAD2 the β-amyrin is not detected anymore and the SAD 2 product (12,13-epoxy-16-hydroxy-β-amyrin, **b**) is detected as a purple spot, consistent with K. Geisler (Osbourn lab) TLC analysis (lane 4, Fig. 5.5.A,), this compound is more polar than β-amyrin. Accumulation of β-amyrin decreases in SAD1/UGT73C10 expressing tissues compare to SAD1 (Fig. 5.5.A, lane 5). The same phenomenon was observed for SAD1/SAD2 product (**b**) in tissues co-expressing SAD1/SAD2/UGT73C10 compare to SAD1/SAD2 (Fig. 5.5.A, lane 6). Those results are supported by GC-MS analysis of extracts from SAD1/SAD2/GFP and SAD1/SAD2/UGT73C10 expressing tissues (Supp. S.8). The absence of SAD2 product in tissues expressing SAD1/SAD2/UGT73C10 even suggest that UGT73C10 entirely consumed 12,13-epoxy-16-hydroxy-β-amyrin. In the meantime, the formation of a new compound (d) is detected when co-expressing UGT73C10 with SAD1 alone or with SAD1 plus SAD2, none of these products (**c** and **d**) are detected when expressing UGT73C10 alone (lanes 1 to 3; Fig. 5.5.B). The

compound **c** has the same retardation factor than purified β-amyrin-3-*O*-glucoside synthetized using recombinant UGT73C10 (Fig. 5.5.B, lanes 2 and 5), suggesting UGT73C10 expressed in *N. benthamiana* is able to use β-amyrin produced by SAD1 to form β-amyrin-3-*O*-glucoside detected *in vitro*. The compound **d** is slightly more polar than **c**, it is likely to be the gluco-conjugated product of SAD2 (Fig. 5.5.B, lane 3). In summary, TLC together with GC-MS analysis suggests that UGT73C10 is able to glucosylate β-amyrin and 12,13-epoxy-16-hydroxy-β-amyrin when transiently co-expressed with appropriate SAD proteins.
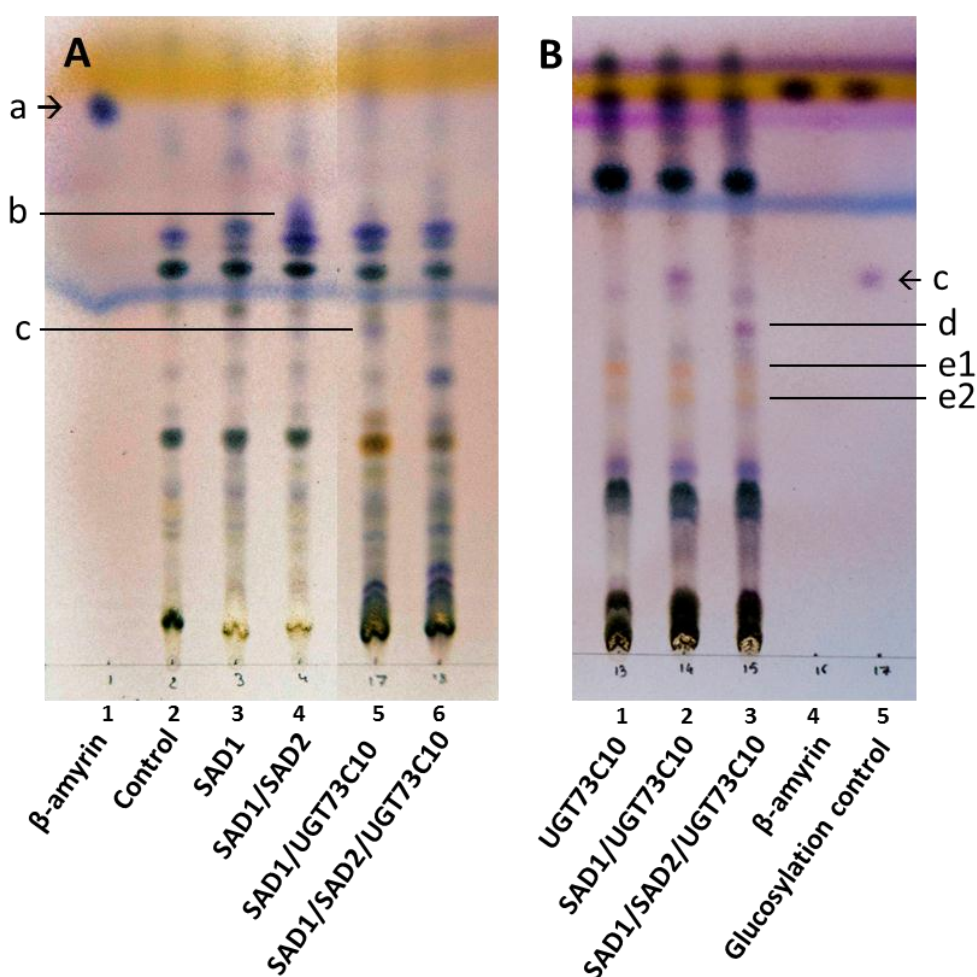
Figure 5.5: Formation and modification of triterpenoids in *N. benthamina*. Thin layer chromatography of *N.benthamina* extracts stained with acetic acid : sulphuric acid : p-anisaldehyde (96:2:2). Methanolic extractions were done with 100mg of agroinfiltrated leaf material from *N.benthamina*, using 75% methanol (**A**) or 40% methanol (**B**) as solvents. **A.** Pure β-amyrin is in lane 1, uninfiltrated tissue is in lane 2, SAD1 only and SAD1/SAD2 expressing tissues in lanes 3 and 4. Tissues co-expressing SAD1 and UGT73C10 in lane 5 and SAD1, SAD2 and UGT73C10 in lane 6. **B.** Tissues expressing UGT73C10 alone (lane 1), with SAD1 (lane2), or with SAD1 and SAD2 (lane3). Pure β-amyrin (lane 4) and reaction product of recombinant UGT73C10 incubated with β-amyrin and UDP-Glc (lane 5). Identified compounds: a- β-amyrin standard, b-12,13-epoxy-16-hydroxy-β-amyrin (glycosylated triterpene standard), c- β-amyrin-3-*O*-glucose, d- new glucosylated triterpenoid, e1 and e2 unknown new products.

It is interesting to notice that additional products (e1 and e2) accumulate in tissues agroinfiltrated with UGT73C10 (figure 5.5.A, lanes 5, 6). These compounds are stained orange by para-anisaldehyde and their accumulation is independent of SAD1 or SAD2 expression. Para-anisaldehyde is a multipurpose staining reagent reacting with nucleophilic chemical groups (van der Heide 1966). Attempt to ascertain which functional groups are present in a molecule based on staining coloration is delicate especially for complex molecules. This compound may be a glucoconjugate of an endogenous compound of *N. benthamiana*.

5.2.2.3 *UGT73C10 attenuates the necrosis observed in SAD1 expressing tissues*

In parallel to accumulation of the triterpenoid compounds **a** and **b**, tissues expressing SAD1 or SAD1/SAD2 developed a yellowing phenotype 6-days after infiltration (pictures 1 and 3, Fig. 5.6). This phenotype quickly evolved to necrosis of the agroinfiltrated tissue. The molecular mechanisms underlying this phenotype are not clearly understood so far and are being investigated by J. Reed (Osbourn lab). The negative effect of β-amyrin accumulation on the intracuticular water barrier may be the cause (Buschhaus and Jetter 2012). Co-expression of SAD1 with UGT73C10 did not lead to leaf yellowing and agroinfiltrated tissues display a wild-type phenotype (picture 2, Fig. 5.6). A similar suppression of the yellowing phenotype is observed in leaves co-infiltrated with SAD1/SAD2 and UGT73C10 expression vectors (picture 4, Fig. 5.6). This observation suggests that the compound responsible for the leaf yellowing (presumably β-amyrin) is detoxified by the activity of UGT73C10. This is consistent with vacuolar storage of many harmful compounds and xenobiotics in plant cells (Bowles et al. 2006; Gachon et al. 2005b; Loutre et al. 2003).
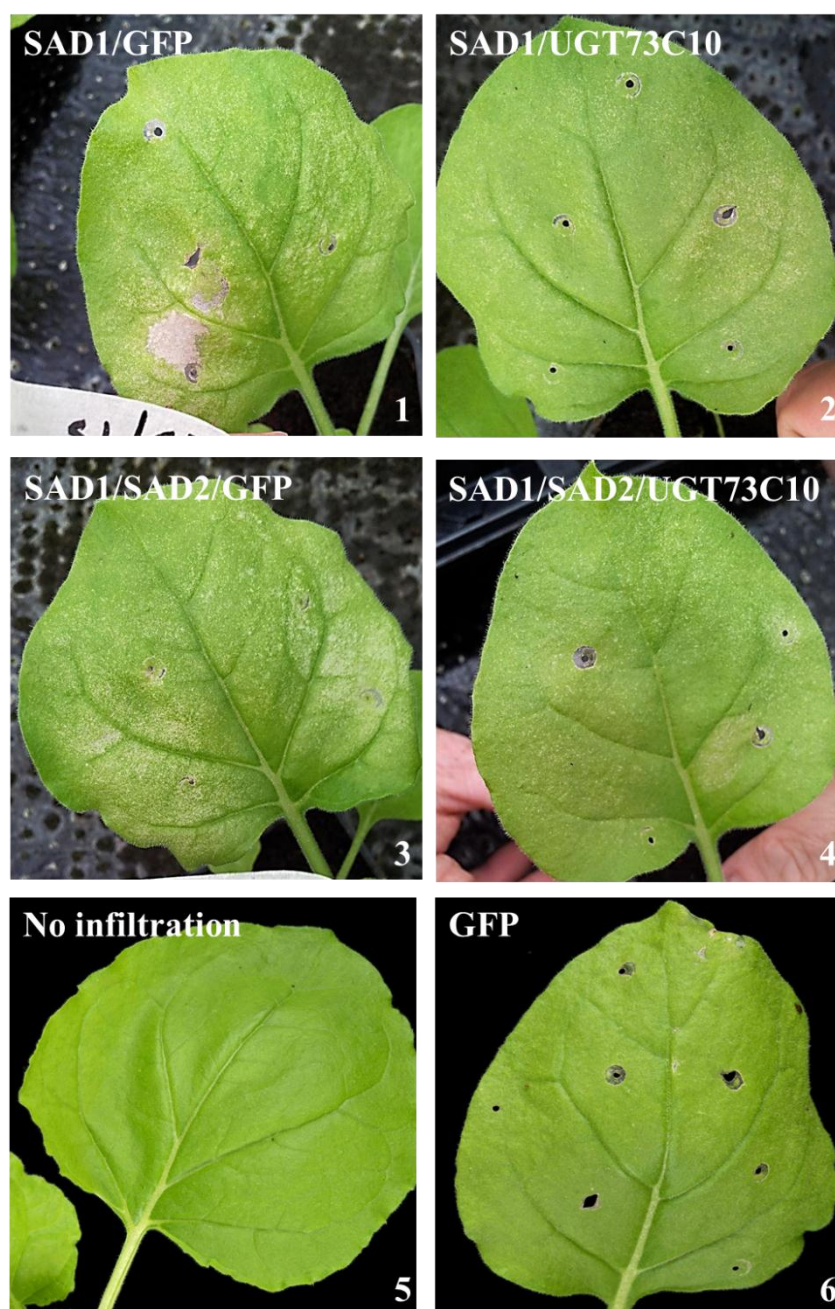
Figure 5.6: Effect of UGT73C10 expression on leaf phenotype related to SAD1 and SAD2 expression. Pictures of *N. benthamiana* leaves co-agroinfiltrated with SAD1/GFP (1), SAD1/UGT73C10 (2), SAD1/SAD2/GFP (3), SAD1/SAD2/UGT73C10 (4) expression vectors (pEAQ-*HT*-CDSs). Controls are non-agroinfiltrated leave (5) and GFP-expressing leave (6). GFP was used as a control in co-expression analysis. These pictures have been taken 4 days after agroinfiltration.

5.2.2.4   *Identification of SAD1-SAD2-UGT73C10 co-expression product*

In order to confirm the accumulation of a glycosylated 12,13-epoxy-16-hydroxy-β-amyrin in *N. benthamiana* agroinfiltrated tissues we develop a bulk purification method (see section 2.2.24) of the compound **d** (Fig. 5.5). Three plants were agroinfiltrated with pEAQ-*HT*-UGT73C10 and pEAQ-*HT*-SAD1-SAD2, 565 mg of dry material was harvested six days after infiltration. Plant material was extracted and enriched in the targeted compound **d** prior to component separation by silica gel column chromatography followed by semi-preparative HPLC monitored by a charged aerosol detector (see section 2.2.24). The total weight of the purified compound was estimated at 250 μg, yielding approx. 0.5 mg/g of dry material (Supp. S.9).

A direct injection of the purified compound was performed using a LTQ-OrbitrapTM mass spectrometer (Thermo Fisher Scientific Inc. see section 2.2.25). The exact masses calculated for the suggested product (12,13-epoxy-16-hydroxy-β-amyrin glucoside; $C_{36}H_{60}O_8$) was m/z 621.4361 $[M+H]^+$ and m/z 643.4186 for the sodium adduct $[M+Na]^+$; these ions were present in the HR-MS spectrum $[M+H]^+$ m/z 621.4364 and $[M+Na]^+$ m/z 643.4179 (respective errors 0.05% and 0.1%; Fig. 5.7.A). The fragment m/z 603.4255 corresponds to the exact calculated mass of the dehydrated compound $[M-H_2O+H]^+$. High resolution MS, together with GC-MS analysis suggests the product of SAD2 is used by UGT73C10 as an acceptor leading to the accumulation of a glucoconjugate of 12,13-epoxy-16-hydroxy-β-amyrin in *N. benthamiana* leaves.

The purified compound was analysed by proton NMR to confirm the overall structure of the predicted product (Fig. 5.7.B). Signals from HC-3, HC-12 and HC-16 are visible in addition to the 6 protons (C$\underline{H}$-OH) of the hexose between 2.4 to 4 ppm range. The signal of CH-1' (4.3 ppm, 1H, d, $J_{1',2'}$=7.76 Hz) is characteristic from a beta anomer of a sugar. Signals from the 8 methyl groups of the triterpene backbone integrating three protons as singlets are localised between 0.5 to 1.4 ppm range. The presence of the eight methyl groups of the β-amyrin backbone is confirmed by HSQC analysis (Supp. S.10).
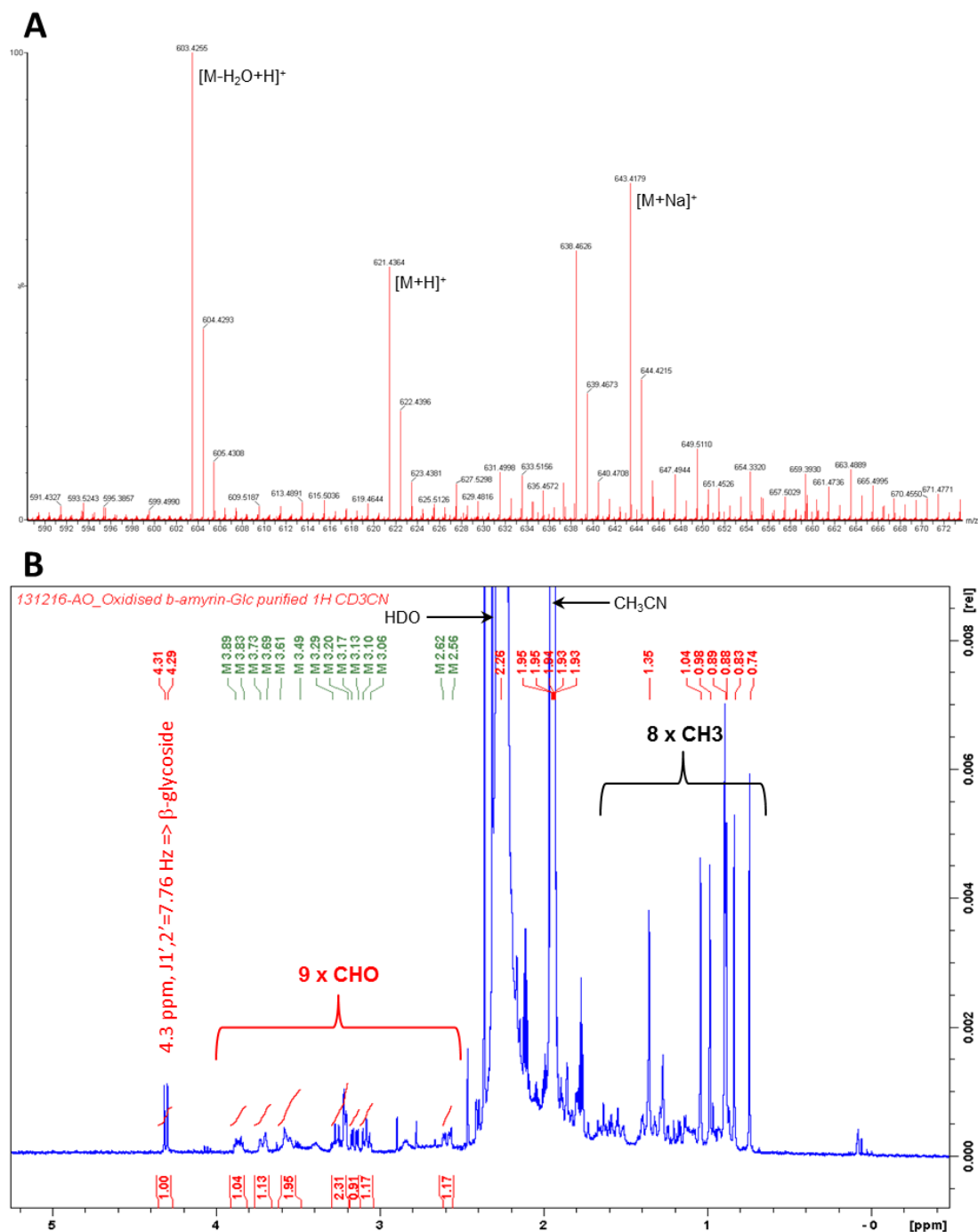
Figure 5.7: Identification of SAD1-SAD2-UGT73C10 co-expression product extracted and purified from agroiniltrated *N. benthamiana*. A. High resolution mass analysis of the purified compound using direct injection onto LTQ-OrbitrapTM mass spectrometer (Thermo Fisher Scientific Inc.). The exact masses of [M+H]$^+$ m/z 621.436, [M+Na]$^+$ m/z 643.418 and [M-H$_2$O+H]$^+$ m/z 603.425 are indicated on the spectrum. **B.** NMR analysis of the purified SAD1-SAD2-UGT73C10 co-expression product. The anomeric proton at position C1' is characteristic of a β anomer (4.3 ppm, 1H, d, 7.76 Hz). In the range of 2.5 to 4 ppm the 6 C<u>H</u>-OH protons of the sugar ring are detected (HC-2', HC -3', HC-4', HC-5', HC-6'a and HC-6'b), plus HC-3, HC-16 and HC-12 from the terpenoid backbone. Protons from the 8 methyl groups of β-amyrin are detected in 0.5 to 1.4 ppm (HC-23, HC-24, HC-25, HC-26, HC-27, HC-28, HC-29 and HC-30).

Altogether, TLC, HR-MS, GC-MS and H_1-NMR all indicates the accumulation of a new compound dependant of UGT73C10, SAD1 and SAD2 co-expression in parallel with consumption of the SAD2 product. In the light of Augustin et al. (2012) our data suggest the production of 12,13-epoxy-16-hydroxy-β-amyrin-3-*O*-β-D-glucose.

3.1.1 Co-expression of oat UGT collection with SAD1 or SAD2

pEAQ-*HT*-UGTs constructs were agroinfiltrated in *N. benthamiana* leaf tissue. UGTs expression vectors were agroinfiltrated alone or with SAD1, or with SAD1 plus SAD2 expression vectors (pEAQ-*HT*-SAD1 and pEAQ-*HT*-SAD1-SAD2). Methanolic extracts from the various co-expression combinations were analysed by TLC. *ρ*-Anisaldehyde-stained TLC plates loaded with methanolic extracts (40% methanol) are presented in figure 5.8. TLC analysis suggests that none of the UGTs from our collection was able to form compounds with similar retardation factors to hypothetical 12,13-epoxy-16-hydroxy-β-amyrin glycoconjugate (Fig. 5.8, compound d). Additionally, β-amyrin 3-*O*-glucoside, β-amyrin 3-*O*-arabinoside or compounds dependent on co-expression of the various UGTs with SAD1 or SAD1 and SAD2 were not detected by TLC analysis.

Interestingly agroinfiltrated leaves expressing some of the *A. strigosa* UGTs were able to produce orange stained compounds similar to the one produced by UGT73C10-expressing tissues (Fig. 5.5, compound **e**). As observed for UGT73C10 the production of these compounds is independent of SAD1 or SAD2 expression. Tissues expressing AsGT23586B accumulates two distinct orange-stained compounds while AsGT16f23, AsGT24i2, AsGT25n16 and AsGT23586B accumulates a single compound (black arrows on figure 5.8). The orange compounds are therefore likely to be due to glycosylation of endogenous *N. benthamiana* metabolites. Isolation and identification of this compound is under investigation.

Comparison of methanolic extracts from *N. benthamiana* tissues agroinfiltrated with GFP (control) or AsGT16f23 expression constructs reveals the accumulation of four new products dependent on AsGT16f23 expression (Fig. 5.9.A). Three of these compounds have a fragmentation pattern consistent with glycosides with a fragment loss of 162 daltons corresponding to a hexose residue (Fig. 5.9.B). None of these compounds were identified; the molecular weights of the

hypothetical agycons are 562 Da (peak Rt 5.5 min), 421 Da (peak Rt 13.9 min) and 405 Da (peak Rt 9.8 min). Those compounds may be large terpenes; the first one (MW 562 Da) may be a modified triterpene or steroid, the two last ones are too small to be triterpenoids but may be modified diterpenes. The higher peak (Rt 12.2 min) is not likely to be a glycoconjugated metabolite and may be due to an indirect effect of glycosyltransferase expression.
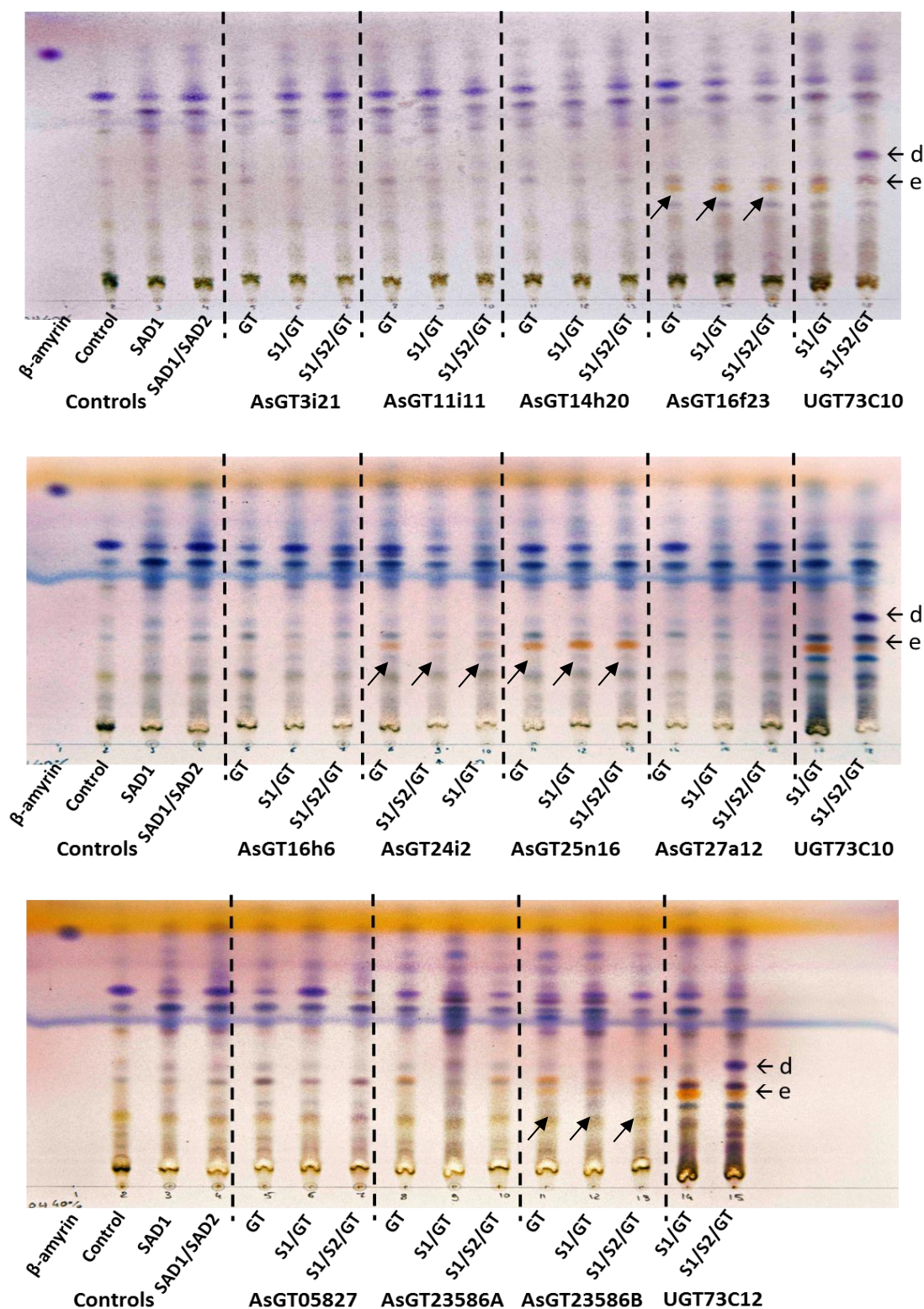
Figure 5.8: Metabolic analysis of *N. benthamina* tissues expressing heterologous *A. strigosa* UGTs alone or with SAD1 and SAD2. Thin layer chromatography of *N. benthamina* extracts stained with acetic acid : sulphuric acid : *p*-anisaldehyde (96:2:2). Methanolic extractions were done on 100mg of agroinfiltrated leaf material from *N. benthamina*, using 40% methanol as solvent. Controls refer to sample detailed in figure 5.5. Each UGT was expressed alone (GT), in combination with SAD1 (S1/GT) and in combination with SAD1 and SAD2 (S1/S2/GT) and loaded in the order mentioned in the figure. Compounds have been labelled as in figure 5.5: d- 12,13-epoxy-16-hydroxy-β-amyrin-3-*O*-β-D-glucose, e- unknown orange-stained product.

Figure 5.9: LC-MS metabolic profiling of *N. benthamiana* tissues agroinfiltrated with AsGT16f23 expression construct reveals accumulation of potential glucosides. A. Full MS spectrum comparison of methanolic extracts from leaf tissues agroinfiltrated with GFP or AsGT16f23 expression constructs. Peak products dependent of AsGT16f23 expression are highlighted with an arrow; biological replication suggests this result is replicable. **B.** MS2 spectrums extracted from three of the four peak products accumulated in AsGT16f23 expressing tissues. Losses of 162 daltons (corresponding to a hexose residue) are labelled on the spectrums.

5.3   Conclusion

In this chapter, *de novo* synthesis of simple triterpenoid saponins in *N. benthamiana* was reported using a combinatorial approach for the co-expression of a member from each enzyme family essential to saponin biosynthesis (OSCs, P450s and UGTs). The structure of the product dependent on co-expression of SAD1, SAD2 and UGT73C10 has been solved as 12,13-epoxy-16-hydroxy-β-amyrin-3-*O*-β-D-glucose.

5.3.1   *Nicotiana benthamiana*, a relevant organism for production of saponins *de novo*

Co-expression of SAD1, SAD2 and UGT73C10 in *N. benthamiana* leaf tissue is a proof of principle that endogenous 2,3-oxidosqualene can be sequentially processed through cyclisation, hydroxylation/epoxidation and glucosylation to create a simple saponin pathway in *N. benthamiana*. Mass spectrometry and proton NMR analysis supported the proposed structure of SAD1/SAD2/UGT73C10 co-expression product as 12,13-epoxy-16-hydroxy-β-amyrin-3-*O*-β-D-glucose. Glycosylation of SAD1 and SAD2 products is also supported by the reduction of the phenotype associated with triterpenoid accumulation in tissues co-expressing SAD1/UGT73C10 or SAD1/SAD2/UGT73C10 (Fig. 5.6); the consumption of the SAD2 in SAD1/SAD2/UGT73C10 expressing tissues was also monitored by GC-MS (Supp. S.8).

Triterpenoids extraction and purification from natural sources is generally a complex task due to low accumulation, often in complex mixtures, of these compounds *in planta*. Synthetic biology approaches constitute a promising way to enhance and diversify the production of triterpenoid saponins for industry (Moses et al. 2013; Sawai and Saito 2011). Very few works have been performed on bioengineering of triterpenoid saponins pathway. OSCs and P450s have been expressed in *Saccharomyces cerevisiae* (Moses et al. 2014; Moses et al. 2013), and metabolic engineering have been done upstream of oxidosqualene in *Nicotiana tabacum* (Chappell et al. 1995; Enfissi et al. 2005; Holmberg et al. 2003; Wu et al. 2012). Here is the first report of *de novo* synthesis of a simple triterpenoid saponin

following heterologous expression of each of the three major enzyme families responsible for cyclisation (OSCs), oxidation (P450s) and glycosylation (UGTs) of saponins. van Herpen et al. (2010) conducted a similar combinatorial biosynthesis approaches on the sesquiterpenoid artemisinic acid (precursor of the antimalarial drug artemisin). LC-MS analysis reveals the accumulation of glycoconjugated artemisinic acid apparently due to endogenous glycosyltransferases of *N. benthamiana*, the resulting glycoside was accumulated to 39.5 mg/kg of fresh weight (Approx. equivalent to 0.4 mg/g according to our measurements). Together with Geisler et al. (2013) report, this work suggests *N. benthamiana* is an excellent platform for saponin triterpenoid engineering. This is suggested by the accumulation of SAD2 product over 1 mg/g of dry weight (dwt) and its entire consumption when UGT73C10 is co-expressed with SAD1 and SAD2. From the small scale purification mentioned in this chapter a yield of 0.5 mg/g dwt have been estimated. The efficiency of the system is likely to be improved significantly considering the impact of upstream MVA pathway engineering on accumulation of β-amyrin, up to 5 folds (J. Reed work, Osbourn lab unpublished data).

5.3.2   A promising system for the discovery of biosynthetic enzymes

pEAQ-*HT*-UGTs constructs were agroinfiltrated in *N. benthamiana* leaf tissue to determine their *A. strigosa* UGTs activity against 12,13-epoxy-16-hydroxy-β-amyrin *in planta* and evaluate their potential implication in avenacin biosynthesis. This approach relies on co-infiltration of SAD1, SAD2 and UGTs expression vectors in order to form glycoconjugates of SAD1/SAD2 product. No potential triterpenoid glucoconjugates were detected in *N. benthamiana* leaf tissue co-infiltrated with *A. strigosa* UGTs and SAD1 or SAD1 and SAD2. Western blot analysis performed with SAD10 suggests active UGT enzymes are spatially and temporally co-expressed with SAD1 and/or SAD2 in harvested tissues. The accumulation of new products related to agroinfiltration of some of the pEAQ-*HT*-UGTs constructs designates the present heterologous system as a valuable tool for investigation of glycosyltransferases activities. Detection methods have to be improved in order to fully exploit this promising heterologous co-expression system. Preliminary LC-MS analysis on tissues expressing AsGT16f23 has shown the accumulation of four new products when compare to control GFP-expressing tissues (Fig. 5.9). Three of these

products have a fragmentation pattern corresponding to glycosylated products (e.g. loss of 162 Da corresponding to hexose residues). These results suggest heterologous expression of UGTs in *N. benthamiana* leads to glycosylation of endogenous compounds and might be used to investigate UGT functions.

5.3.3   General conclusion

In this chapter, the co-expression of members of the three major enzyme families involved in saponin biosynthesis (OSCs, P450s and UGTs) leads to the production of a novel saponin scaffold in *N. benthamiana* leaves. This work is paving the way for future design and engineering of more complex saponins structures by synthetic biology approaches. This system has been used to explore *A. strigosa* UGT activity against an intermediate of the avenacin pathway.

Chapter 6 – Preliminary study of the catalytic activities of *A. strigosa* UGTs towards a subset of potential acceptors

6.1 Introduction

Plant UGTs are promiscuous enzymes that catalyse the formation of glycoconjugates with a wide range of acceptors *in vitro*. UGTs generally show regiospecific activity, catalysing sugar transfer to the same position of related substrates (Bowles et al. 2006; Vogt and Jones 2000). The promiscuity of plant UGTs makes them attractive candidates in the search for biocatalysts that are able to form specific isomers of a given glycoside for biotechnology applications, including a health promoting agent for the food industry, or pharmaceuticals with enhanced solubility and stability. This approach is particularly attractive when compared to chemical synthesis of glycosides of complex natural products, a procedure that depends on availability of the required aglycone, synthesis of the saccharide, and addition and removal of protecting groups (Yu et al. 2012).

Bowles and co-workers have reported the expression and systematic functional analysis of recombinant UGTs from *A. thaliana* (Caputi et al. 2008; Lim et al. 2003; Lim et al. 2002; Weis et al. 2006). Their work greatly contributes to the general understanding of functional evolution of plant UGTs. Some of the *A. thaliana* UGTs were used as whole-cell regioselective biocatalysts for the formation of glycoconjugates of biotechnological interest, like the pharmaceuticals daidzein or *trans*-resveratrol glucosides (Lim et al. 2004; Lim and Bowles 2004; Weis et al. 2006). The collection of active recombinant UGTs obtained from *A. strigosa* roots therefore has considerable potential for the generation of novel glycosides with potential applications in the food, cosmetics and pharmaceutical industry.

6.1.1 The roles of flavonoids and their glycosides

Flavonoids represent a large class of secondary metabolites with more than 10.000 structures reported to date (Buer et al. 2010; Liu et al. 2013). These molecules, which are produced from the phenylpropanoid pathway, are ubiquitous in the plant kingdom. Flavonoids have been reported to have essential functions in defence against various pathogens, including nematodes, fungi and bacteria (Treutter

2005). They also provide protection again UV radiation and have important roles in attracting pollinators by serving as pigments in flowers (Nishihara and Nakatsuka 2011; Tohge et al. 2011). In addition to these ecological roles, flavonoids have also been shown to be involved in regulation of auxin transport and levels of reactive oxygen species (Buer et al. 2010; Weston and Mathesius 2013).

Flavonoids are of particular importance in roots and are often found in root exudates in the rhizosphere, where they may protect the plant against soil-borne pathogens (Buer et al. 2010; Weston and Mathesius 2013). They are also essential players in symbiotic interactions, where they act as chemo-attractants for symbionts such as nodule-forming rhizobial bacteria and arbuscular mycorrhizal fungi (Oldroyd 2013). Some flavonoids also have roles in allelopathy; flavonoids from rice and barley have been associated with the allelopathic properties of these species, along with other chemicals (Weston and Mathesius 2013).

Flavonoids have also attracted considerable interest due to their health promoting effects in humans. Various studies have clearly established a relationship between flavonoid intake and reduced risk of heart disease (Heim et al. 2002). Food containing high levels of flavonoids, in particular antioxidant anthocyanins (such as found in blackcurrants or bioengineered purple tomatoes), has been shown to protect against heart diseases or cancer, (Gopalan et al. 2012; Klee 2013).

Within the plant cell, flavonoid glycosides are normally stored in the vacuole (Weston and Mathesius 2013). Glycosylated flavonoids are believed to have little biological activity and are regarded as storage forms of these natural products. Glycosylation of flavonoids may affect their activity and also their absorption by the human body. Generally, aglycones are only absorbed after hydrolysis of the sugar moiety in the small intestine by human β-glucosidases or in the colon by the bacterial flora. Nevertheless, flavonoid glycosides may be absorbed intact in some cases (Cao and Prior 1999; Heim et al. 2002). Consequently the major effect of glycosylation is to delay the absorbtion of the flavonoid until it reaches the intestine. Therefore understanding glycosylation of flavonoids in cereals is of great interest for engineering health-promoting foods and for the generation of enhanced pharmaceuticals.

UGT73 enzymes (group D) constitute the majority of our collection of recombinant enzymes from oat. This family also contains the clade IIIa group of flavonoid glycosyltransferases that have been reported from other plant species (Fig. 2.2).

6.1.2   <u>Sesquiterpenoids: natural products of interest to industry</u>

Sesquiterpenoids are natural products that occur in many plant species and also in other organisms (Cao and Prior 1999; Hemmerlin et al. 2012). They are C15 terpenoids that are produced primarily by the MVA pathway (see section 1.3.1), although the MEP pathway may also contribute to the synthesis of some sesquiterpenes (Yu and Utsumi 2009). Sesquiterpenoids participate in many biological functions in plants. The phytohormone abscisic acid (ABA) is a sesquiterpenoid that is involved in essential processes such as seed dormancy, stomatal closure, vegetative growth and environmental stress responses (Piotrowska and Bajguz 2011). Some sesquiterpenoids have roles in plant defence - for example capsidiol, a phytoalexin produced by Solanaceaous plants in response to oomycete pathogens (De Marino et al. 2006; Threlfall and Whitehead 1988; Yu 1995). Volatile triterpenoids contribute to floral scents (Baez et al. 2012; Custodio et al. 2006; Theis and Raguso 2005) and are also emitted to attract predators of herbivores that feed on plants (Turlings and Tumlinson 1992; Turlings et al. 1990). In cereals, very little is known about occurrence and functions of sequiterpenoids, although a family of antifungal sesquiterpenoid phytoalexins known as zealexins has recently been identified in maize (Huffaker et al. 2011).

Sesquiterpenoids are bioactive compounds that have been used by humans since ancient times for various applications. Many of these compounds have pharmaceutical properties. Artemisinin, from *Artemisia annua,* is one of the most famous examples, due to its antimalarial activity (Lee 2007). Artemisinin is also a potent antitumor agent, as are other sequiterpene lactones (Ghantous et al. 2010). Zerumbone and farnesol are also very promising chemicals for cancer treatment (Kuete and Efferth 2013; Prasannan et al. 2012). Sesquiterpenes are generally apolar compounds and may be volatiles, those properties are limiting when it comes to potential pharmaceutical applications. Derivatization and particularly glycosylation might modify those properties.

Sesquiterpenoid glycosides have been identified from diverse plant species, and there are many reports of the detection of these compounds in roots (Kitajima et al. 2003; Wang et al. 2013c; Zhuang et al. 2013). From these reports it transpires that if glycosylated, sesquiterpenes usually have a simplistic glycosylation pattern consisting of monosaccharides or short sugar chains. Interestingly, the analysis of root metabolites composition through the Poaceae family revealed the accumulation of two sesquiterpene diglycosides in *Avena sativa* induced by an arbuscular mycorrhizal fungus (Maier et al. 1997). The sugar moities of both of these compounds are similar and consists of a β-D-glucose-2-β-D-glucuronate; the sesquiterpene moiety consists of blumenol A or B. To my knowledge, with the exception of ABA, very little is known about sesquiterpenoid glycosylation *in vivo*. Glucosylated ABA is known to be an inactivated form of the phytohormone, and may also be used for transport (Piotrowska and Bajguz 2011). Efforts have been made to identify ABA glycosyltransferases (Lim et al. 2005; Suzuki et al. 2007; Xu et al. 2002) and the enzyme UGT71B6 has been shown to increase ABA-Glc levels when overexpressed in *A. thaliana* (Priest et al. 2006; Priest et al. 2005). A number of sequiterpenoids have been tested as potential acceptor for the suite of recombinant *A. thaliana* UGTs cloned and expressed by Bowles and co-workers (Caputi et al. 2008). Glucosylation of farnesol was performed by UGTs from group E, D and G; ester-forming enzymes from group L were exclusively involved in glycosylation of artemisinic acid and retinoic acid (Fig. 6.1).
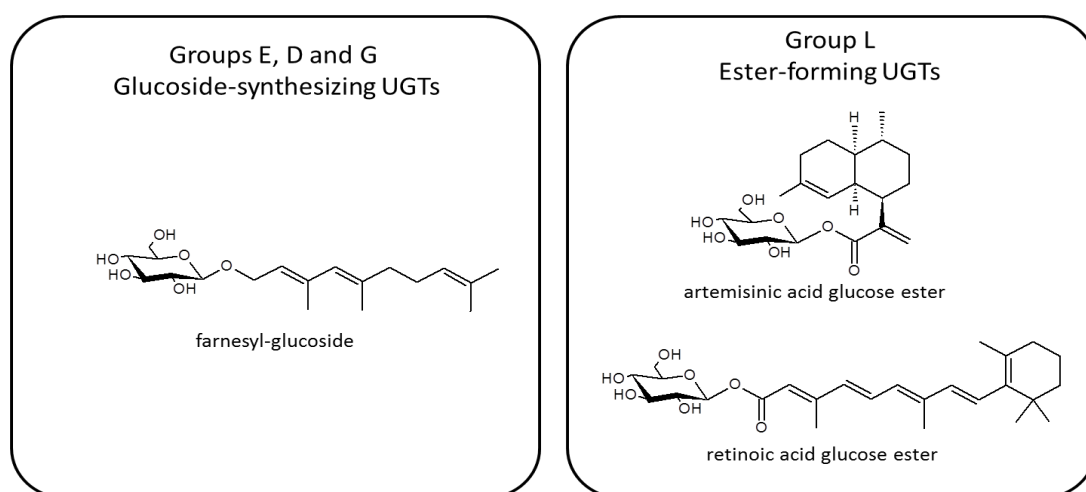
Figure 6.1: Glucosylation of sesquiterpenoids by *A. thaliana* UGTs from various groups (Caputi et al. 2008).

6.1.3   Aims

The large collection of recombinant UGTs that we have assembled for *A. strigosa* is unprecedented for monocot species, to our knowledge. In order to investigate the functions of monocot UGTs and to put these functions into a phylogenetic framework, this preliminary study evaluated the ability of these enzymes to glycosylate a suite of terpenoids and flavonoids. These compounds were chosen firstly because they were potential acceptors for UGT73 enzymes, which represent the majority of the recombinant *A. strigosa* UGT enzymes within our collection, and secondly due to interest in the glycoconjugates of these compounds for food and medical applications.

6.2   Result and discussion

6.2.1   Identification of acceptors of *A. strigosa* UGTs using radioassays

6.2.1.1   *Development of a rapid radioassay to screen for glycosyltransferase activity*

Radioassays are commonly used for enzymatic studies of glycosyltransferases (Augustin et al. 2012; Caputi et al. 2008; Shibuya et al. 2010). Radioactive sugar donors have been used to enhance the sensitivity of glycosyltransferase assays with chemical species that are not detectable by spectroscopic methods. Radioassay experiments with UGTs normally involve detection of the incorporation of radioactive sugar molecules into glycosylated products after TLC separation of the chemical species within the assay. This approach involves a time-consuming protocol unfitted for acceptors screening over a large collection of glycosyltransferases. Here I used a simplified protocol that allowed rapid screening of the glucosylation activity of multiple *A. strigosa* enzymes towards a set of potential acceptors.

Previously I. Ivanova in the laboratory of Rob Field at the John Innes Centre had developed a streamlined radioassay to detect galactosyltransferase activity towards different potential acceptors. This protocol utilises UDP-[1-^{3}H]Gal and relies on the fact that negatively charged sugar nucleotide is retained by ion exchange resin, so allowing chromatographic separation of the unreacted radioactive sugar donor from the glycosylated products. In these assays, detection of radioactivity in the unbound fraction indicates transfer of the radioactive sugar from the donor to the acceptor.

This radioassay protocol was adapted and simplified for use as a screening method for identification of acceptors that are glucosylated by recombinant *A. strigosa* UGTs. In my experiments radioactive UDP-α-D-[6-^{3}H]glucose was used as sugar donor because the experiments with the generic acceptor TCP indicated that the *A. strigosa* UGTs used UDP-α-D-glucose as their preferred sugar donor (see section 4.2.3.3). Assays were carried out as described in section 2.2.14. The ion-exchange resin was placed directly in the tube containing the incubated reaction mixture. The radioactivity remaining in solution after adding the resin was measured

by scintillation counting in order to evaluate transfer of the radioactive glucose onto the potential acceptor molecules (Fig. 6.2).
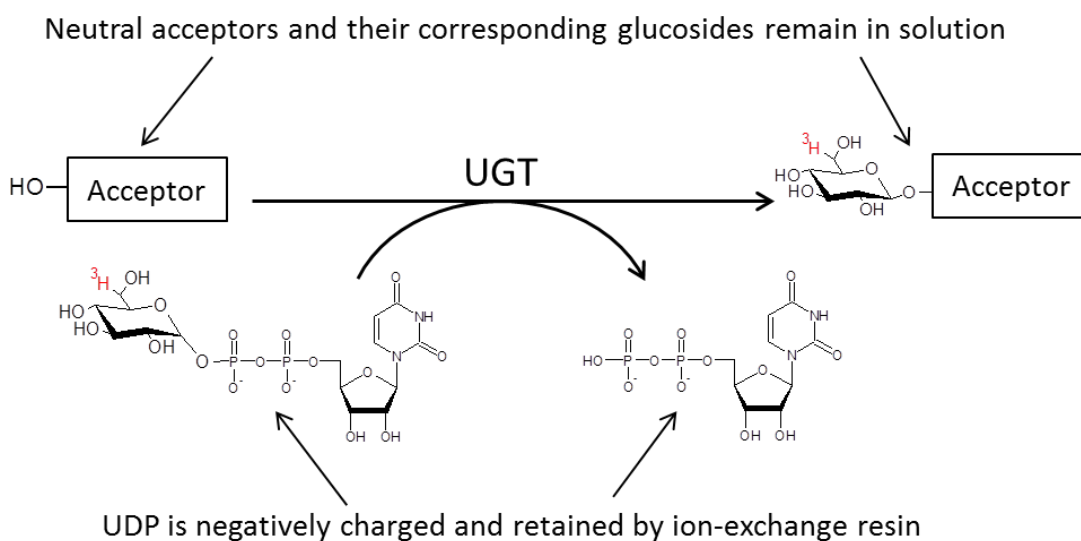
Figure 6.2: Illustration of the basic principles behind the screening method developed based on the protocol of I. Ivanova.

The absence of column chromatography resulted in high background in the preliminary analysis performed with SAD10 and TCP (see section 0). The signal detected in the negative controls (no enzyme, no acceptor) was approximately 25% of the total radioactivity of the assay (Fig. 6.3.A). This background was reduced to less than 5% after optimization of the radioassay (Fig. 6.3.B). The signal to noise ratio was greatly improved when using UDP-Glc as the limiting substrate (50 µM UDP-Glc and 100 µM acceptor). In those conditions, the proportion of UDP-Glc substrate converted in UDP is enhanced; consequently, the radioactive signal in solution due to glucosyl transfer increased compare to the entire radioactivity of the assay. Details of the protocol were also modified to optimize binding of the unreacted sugar donor to the resin and washing steps. Binding of reaction products was done with a larger volume of resin (500 µl instead of 100 µl) and a smaller volume of sample (dilution of reaction products was done after five minutes of incubation with the resin) to improve binding capacity and increase interaction between unreacted substrate and resin. Larger volumes were used in washing steps (1 ml instead of 500 µl) to optimize the amount of unbound material collected and avoid their contamination by resin particles.
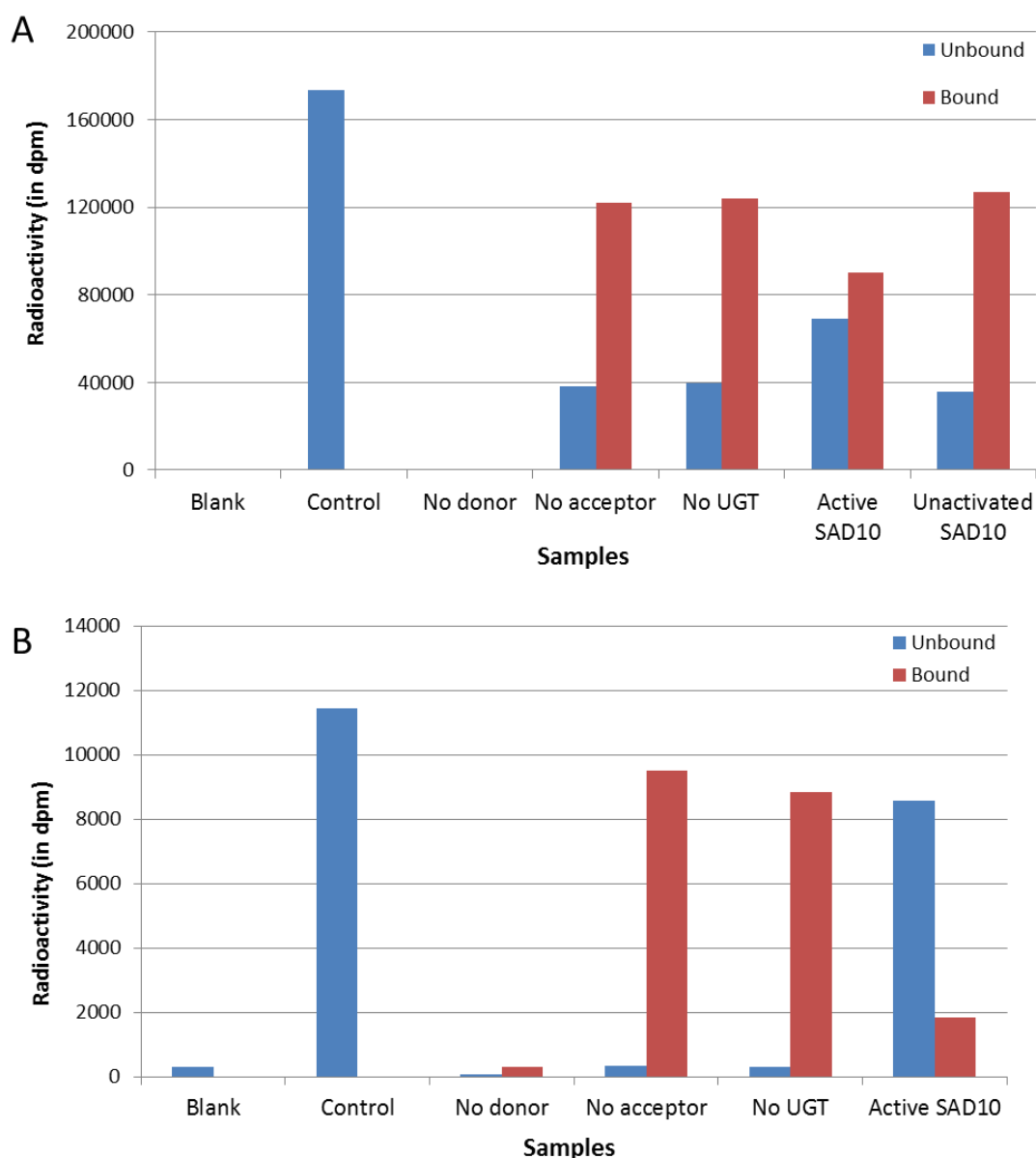
Figure 6.3: Optimisation of a rapid radioassay for the analysis of glucosyltransferase activity. SAD10 glucosylation reaction of TCP was used to assess the optimisation of radioassay conditions. The red bars show the amount of radioactivity in solution after reactions performed with UDP-Glc [6-^{3}H] and 100 μM TCP. The blue bars show the amount of radioactivity in solution after the reaction. Blank and control are scintillation readings with water and with the entire radioactive material used in each assay**. A.** The initial activity assay of SAD10 shows a background noise of approximately 40000 dpm over 170000 dpm (approx. 25% of the total radioactivity of the assay). **B.** Following optimisation of the radioassay protocol the background noise was reduced to 300dpm (less than 5% of the total radioactivity of the assay).

To test the efficacy of the radioassay method for detection of glycosyltransferase activity, reactions with each of the recombinant UGTs were performed using UDP-[^{3}H]-Glc and TCP as substrates (Fig. 6.4). The results of these radioassays are consistent with previous HPLC analysis of the glucosylation reactions performed with TCP (see section 4.2.3). This analysis was duplicated and similar results were obtained, demonstrating the reproducibility of the implemented method (note the error bars reported on figure 6.4). The highest activities were observed for SAD10, AsGT16f23, AsGT24i2, AsGT05827, AsGT23586A and AsGT23586B, with radioactivity in solution from 8500 to 15300 dpm (representing 28.5% to 51% conversion of TCP). This is consistent with the results obtained in section 4.2.3, those 6 enzymes being the more active towards TCP. The radioactivity detected in solution was extremely high for AsGT21p16 both in the presence and absence of acceptor (66% and 49% of the total radioactivity respectively). The reason for this is unclear, although it could be due to hydrolysis of UDP-Glc by a component from the enzyme solution leading to release of free glucose in solution. Control TLC was not explored due to limited amount of time. It is known that misfolded proteins may have modified activities, and misfolding of glycosyltransferases might lead to hydrolysis activity. Therefore potential misfolding of AsGT21p16 may result in hydrolysis of the glycosidic bond of the sugar donor. Another explanation might be that a component of the system is used as an acceptor by AsGT21p16 (e.g. ethanol, included to solubilise acceptor, or traces of imidazole). Release of neutral radioactivity in the absence of an acceptor was also detectable to a lesser degree for AsGT27f7 and AsGT05827 (9% and 16% of the total radioactivity, respectively).
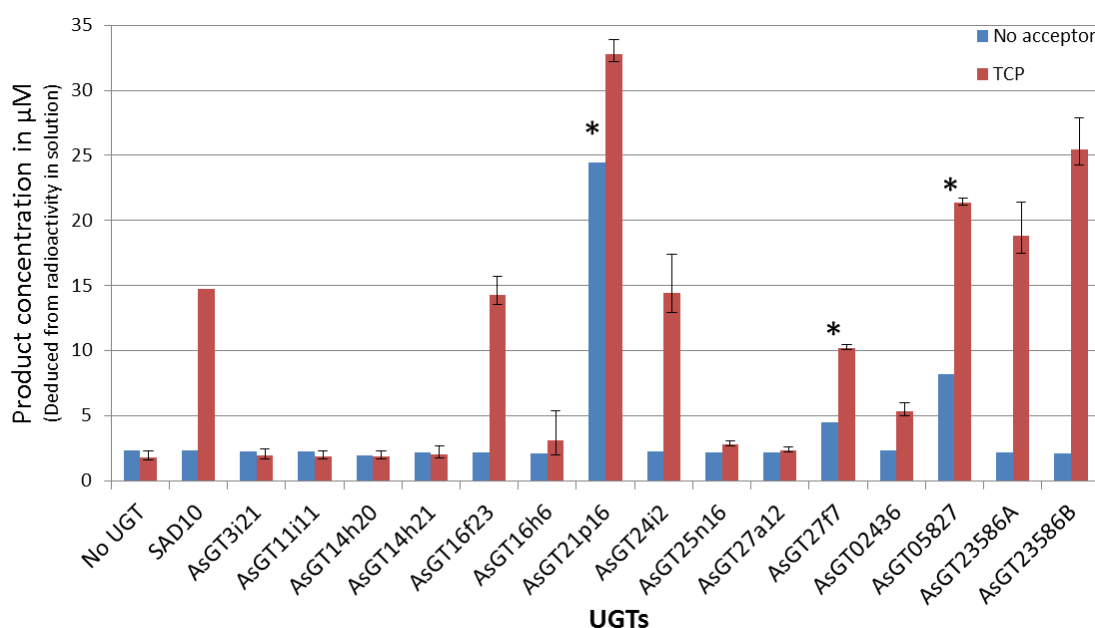
Figure 6.4: Monitoring glucosylation using the radioassay: proof-of-principle using trichlorophenol as the acceptor. The radioassay protocol was tested with the collection of recombinant *A. strigosa* UGTs with TCP as an acceptor. The red bars show the concentration of product formed after reactions performed with 50 μM UDP-Glc [6-^{3}H] and 100 μM TCP obtained from two independent experimental sets. The blue bars show negative control in the absence of potential acceptor. Product concentration was deduced from the radioactivity measured in solution after reactions. Asterisks show values inconsistent with results of section 4.2.3 (discussed in text).

6.2.1.2 *Activity of selected A. strigosa UGTs towards a suite of potential flavonoid and terpenoid acceptors*

Radioassays were performed with a suite of flavonoid and terpenoid acceptors that may represent potential acceptors for enzymes from UGT73 family (see section 3.2.4). As described before, the functions of UGT73 enzymes in dicots appear to be primarily glycosylation of triterpenoids and related compounds (such as sterols or steroidal alkaloids), and flavonoids.

The flavonols kaempferol (1) and quercetin (2) were included as potential acceptors in these assays (Fig. 6.5). These compounds are widely represented in nature and have hydroxyl groups at positions C-3, C-5 and C-7, which are known to be the most common sites of glycosylation for flavonols (Heim et al. 2002). The flavone, tricin (5,7,4'-trihydroxy-3',5'-dimethoxyflavone) (3) is found in various cereals such as oat, rice, barley, wheat and maize (Moheb et al. 2013). Recent reports have shown anticancer activities of tricin (Cai et al. 2009; Oyama et al. 2009). Tricin

possesses hydroxyl groups in position C-7, C-5 and C-4', which are therefore potential sites of glycosylation (Fig. 6.5).
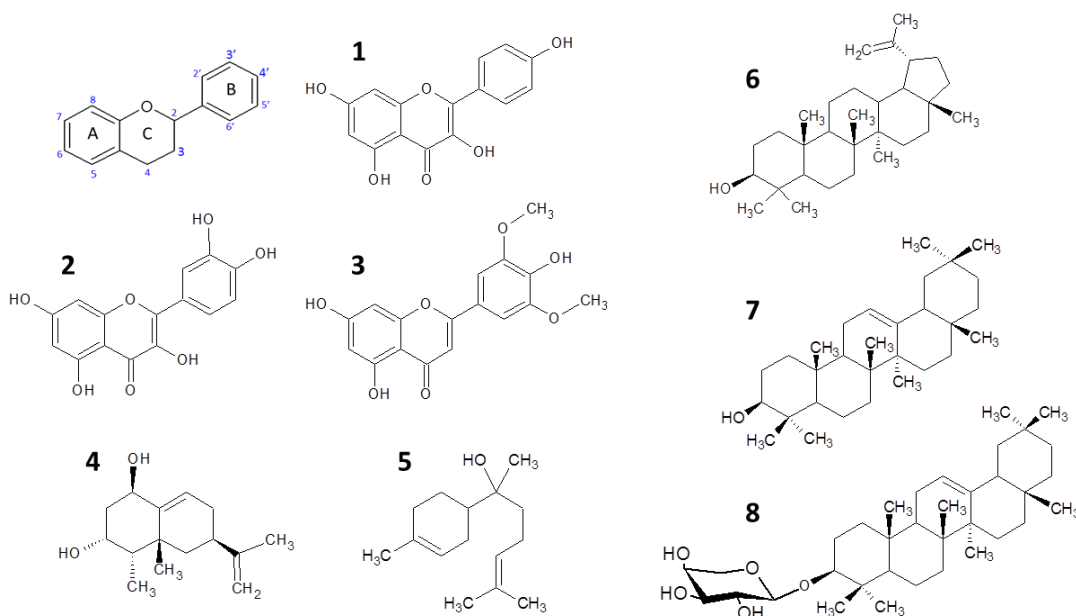
Figure 6.5: Flavonoid and terpenoid used as acceptors in radioassays. Flavonoids: kaempferol (1), quercetin (2) and tricin (3). Sesquiterpenoids: capsidiol (4) and α-(-)-bisabolol (5). Triterpenoids: lupeol (6), β-amyrin (7) and β-amyrin-3-*O*-arabinose.

The sesquiterpenes used included α-bisabolol (5), a volatile sesquiterpene that is produced by several plant species and that has been shown to have various pharmaceutical properties (Darra et al. 2008), including apoptotic effects on glioma cells (Cavalieri et al. 2004). Capsidiol (4) was also included. In addition to its phytoalexin properties (Literakova et al. 2010), capsidiol has been shown to have bacteriostatic properties against *Helicobacter pylori*, a gastrointestinal bacterium linked to the development of stomach ulcers and cancer (De Marino et al. 2006). Capsidiol and α-bisabolol have two and one hydroxyl groups, respectively, and are potential acceptors for *O*-glycosyltransferases (Fig. 6.5). The absence of blumenol A or B from commercial sources prevents the corresponding activity assays.

Triterpenoid acceptors were used to complement the results obtained with cold substrates. Acceptors included were the non-acidic triterpenes available to us: β-amyrin (6), lupeol (7) and β-amyrin 3-*O*-α-arabinopyranoside (8) (Fig. 6.5). Hederagenin (pKa 4.63) and oleanolic acid (pKa 2.52) are negatively charged in conditions used (pH 7.5) and will be retained by the ion exchange resin.

Eight *A. strigosa* UGT enzymes were selected for evaluation against different acceptors based on their higher glucosylation activity over TCP (sections 4.2.3 and 5.1.1.1). This selection included one member of each phylogenetic cluster of UGT73 represented in *A. strigosa* (M1, M2, M3 and M5; section 3.2.4). These were SAD10, AsGT14h20, AsGT16f23, AsGT24i2, AsGT02436, AsGT05827, AsGT23586A and AsGT23586B.

Table 6.1: Relative activity observed in radioassays

Acceptors	Relative activity (%)*		
	AsGT24i2	AsGT02436	AsGT05827
TCP	100	100	100
Quercetin	62	n.d	n.d
Kaempferol	85	350	84
Tricin	37	86	95
Capsidiol	4	n.d	n.d
α-Bisabolol	n.d	n.d	29
Lupeol	n.d	n.d	n.d
β-Amyrin	n.d	n.d	n.d
β-Amyrin-Ara	9	21	n.d

* Relative activity has been deduced from radioassays, the experimental conditions are detailed in section 2.2.14.
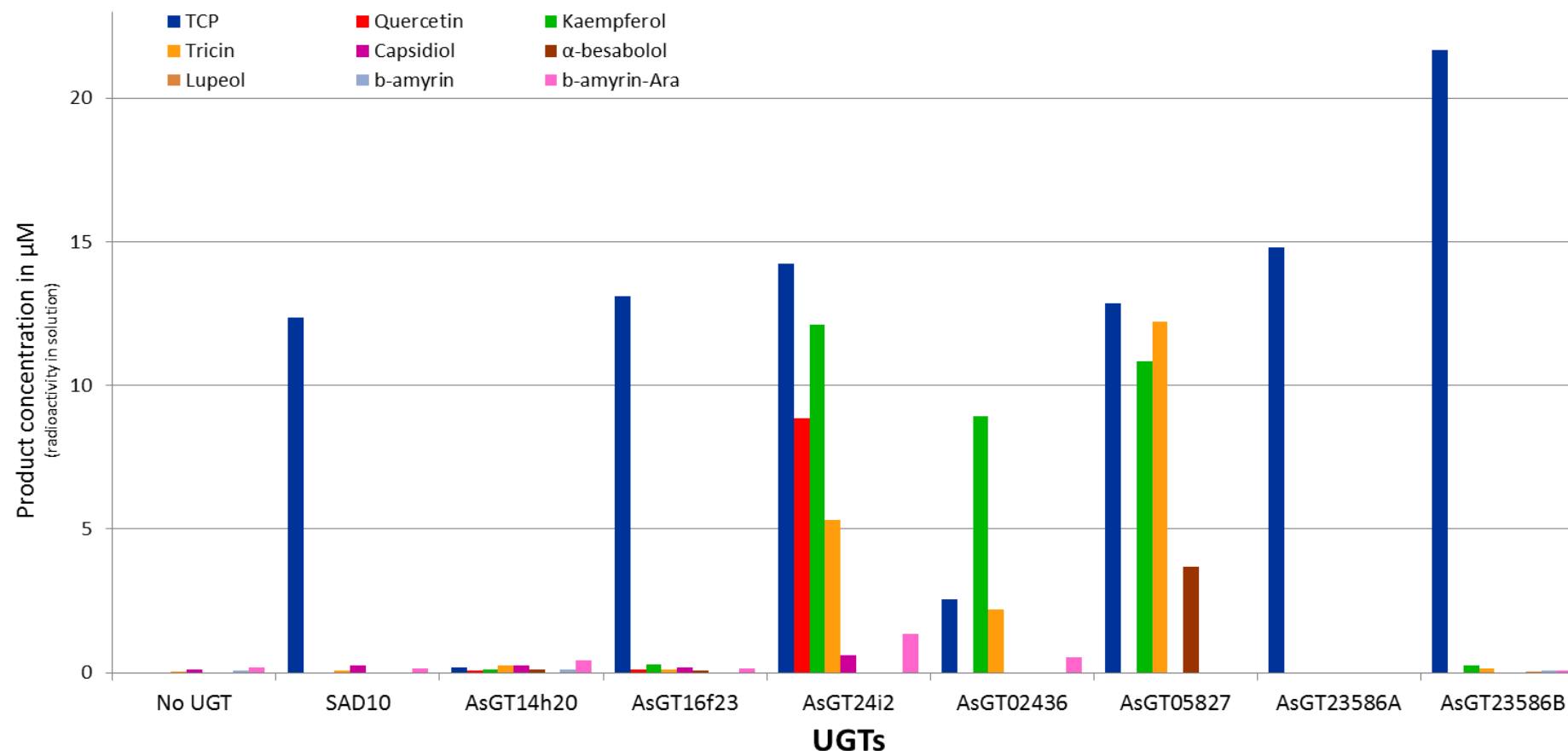
Figure 6.6: Radioassays of 8 selected *A. strigosa* UGTs over a set of acceptors. Glucosylation radioassays were performed with recombinant UGTs using UDP-β-D-[^{3}H]. glucose. Reaction mixtures were composed of 100 mM TRIS-HCl pH 7.5, 50 µM UDP-Glc, 200 µM acceptor and 2 µg of recombinant UGT. Flavonoids and terpenoids were used as potential acceptors. Samples were incubated 12 hours at 35°C. Bars represent the concentration of product formed after reactions for each sample. Product concentration was deduced from the radioactivity measured in solution after subtraction of the background noise (radioactivity detected without acceptor). These results were obtained from a single experiment. Further details about calculations used to generate this figure are presented in section 2.2.14.

The radioactivity in solution detected for each reaction is presented as a diagram in figure 6.6 (after subtraction of the background noise); these results were obtained from a single experiment. The major activities detected in radioassays have been converted into relative activities shown in table 6.1. The highest activities were obtained with flavonoids, in particular with AsGT24i2, AsGT02436 and AsGT05827. AsGT24i2 was able to convert efficiently the three flavonoids tested under the experimental conditions used; conversion of TCP and kaempferol are very similar for this enzyme (Kaempferol conversion represents 85% of TCP conversion), slightly lower conversions were observed for quercetin and tricin (62% and 37% of TCP conversion, respectively). AsGT02436 and AsGT05827 did not show any activity with kaempferol, suggesting the presence of 3'-OH and/or 3-OH affects the activity of these enzymes (Fig. 6.5). Under the conditions used, AsGT02436 shows a clear preference for kaempferol compared to tricin and TCP (kaempferol conversion represents 350% of TCP conversion). AsGT05827 converts kaempferol, tricin and TCP to a similar extent (conversions of kaempferol and tricin represent 84% and 95% of TCP conversion); nevertheless, these results have to be considered carefully due to the activity observed in the absence of acceptor substrate (see the previous section 6.2.1.2). Assays with potential terpenoid acceptors generally resulted in weaker radioactive signals that were only just above background levels. Nevertheless, the activity of AsGT05827 towards α-bisabolol represents 29% of its activity towards TCP. Additionally, the activity displayed by AsGT24i2 towards β-amyrin-3-*O*-Ara represents 21% of the one detected with TCP.

As mentioned earlier, these preliminary results have been obtained from a single experiment and glycoside formation needs to be confirmed to support activities detected using the radioassay.

6.2.2   Further investigation by TLC analysis

The above radioassays are indicative of glucosyltransferase activities towards several acceptors. TLC analysis was then carried out in order to support the formation of flavonoid or terpenoid glucosides.
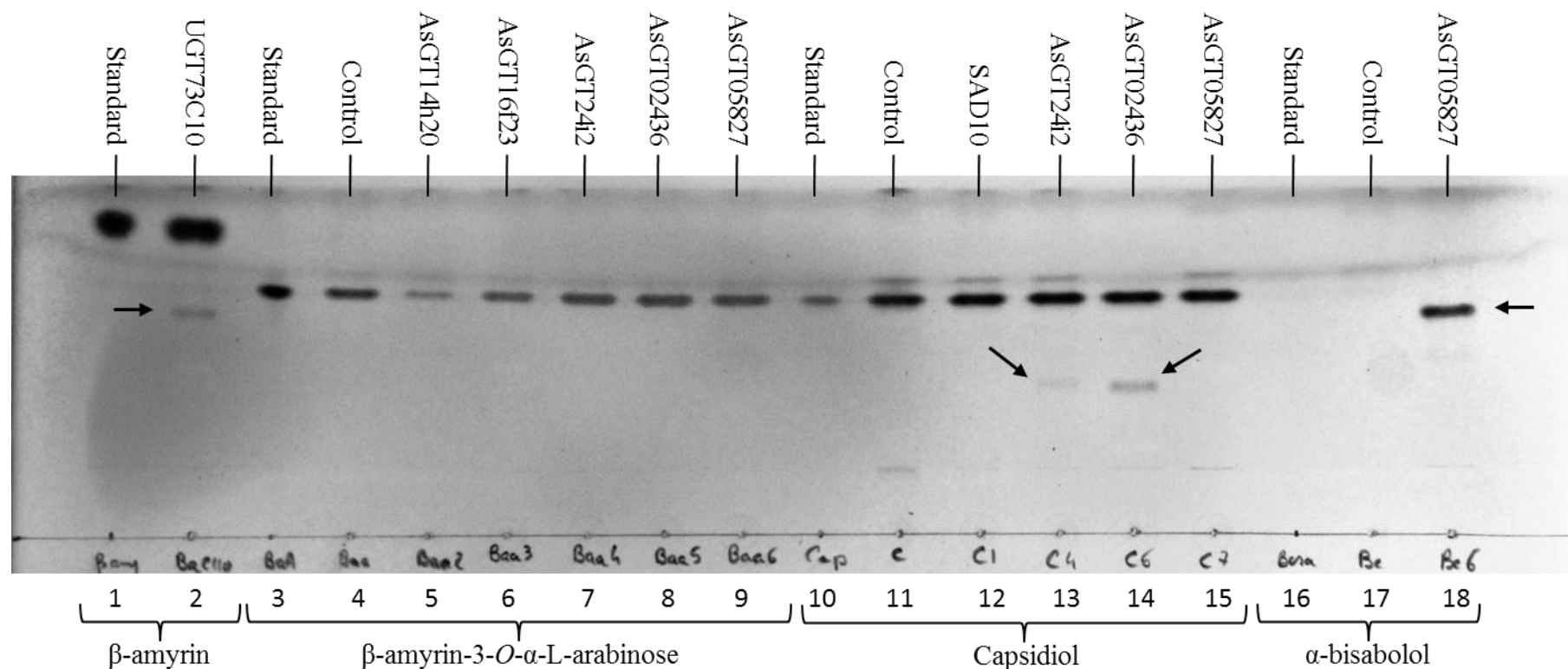
Figure 6.7: Activities of *A. strigosa* recombinant UGTs over terpenoids. Glucosylation assays were performed with recombinant UGTs using UDP-Glc and a set of terpenoid acceptors. Reaction were performed at 35°C for 12 hours using 1mM of UDP-Glc and 200 µM of acceptor. Glucosylated products were extracted in ethyl acetate. Lane 1, β-amyrin standard; ;lane 2; reaction UGT73C10 with β-amyrin; lane 3, β-amyrin-3-*O*-arabinose standard; lane 4, control β-amyrin-3-*O*-arabinose in reaction conditions; lane 5, reaction AsGT14h20 with β-amyrin-3-*O*-arabinose; lane 6, reaction AsGT16f23 with β-amyrin-3-*O*-arabinose; lane 7, reaction AsGT24i2 with β-amyrin-3-*O*-arabinose; lane 8, reaction AsGT02436 with β-amyrin-3-*O*-arabinose; lane 9, reaction AsGT05827 with β-amyrin-3-*O*-arabinose; lane 10, capsidiol standard; lane 11, control capsidiol in reaction conditions; lane 12, reaction SAD10 with capsidiol; lane 13, reaction AsGT24i2 with capsidiol; ; lane 14, reaction AsGT05827 with capsidiol; lane 15, reaction AsGT02436A with capsidiol; lane 16, α-bisabolol standard; lane 17, control α-bisabolol in reaction conditions; lane 18, reaction AsGT05827 with α-bisabolol. Arrows are pointing toward the reaction products detected. It is important to notice that α-bisabolol is a volatile compound and has been lost during TLC processing (TLC plate heater is used to warm the plate at 45°C when loading), only the likely glycoconjugated form is detected here.
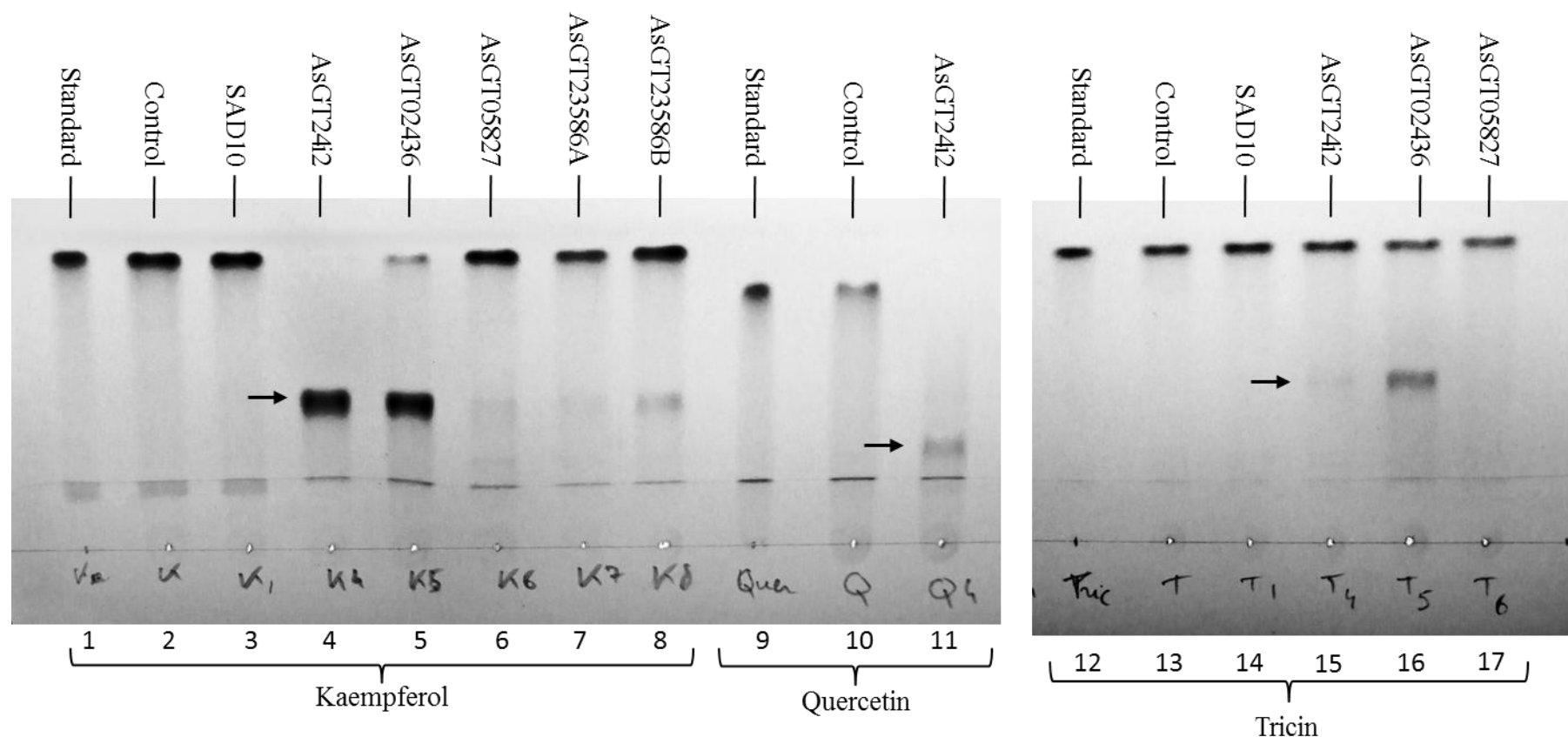
Figure 6.8: Activities of *A. strigosa* recombinant UGTs over flavonoids. Glucosylation assays were performed with recombinant UGTs using UDP-Glc and a set of flavonoid acceptors. Reaction were performed at 35°C for 12 hours using 1mM of UDP-Glc and 200 μM of flavonoid. . Glucosylated products were extracted in ethyl acetate. Lane 1, kaempferol standard; ;lane 2; control kaempferol in reaction conditions; lane 3, reaction SAD10 with kaempferol; lane 4, reaction AsGT24i2 with kaempferol; lane 5, reaction AsGT02436 with kaempferol; lane 6, reaction AsGT05827 with kaempferol; lane 7, reaction AsGT23586A with kaempferol; lane 8, reaction AsGT23586B with kaempferol; lane 9, quercetin standard; lane 10; control quercetin in reaction conditions; lane 11, reaction AsGT24i2 with quercetin; lane 12, tricin standard; lane 13; control tricin in reaction conditions; lane 14, reaction SAD10 with tricin; lane 15, reaction AsGT24i2 with tricin; lane 16, reaction AsGT02436 with tricin; lane 17, reaction AsGT05827 with tricin.

The reactions from which highest activities have been detected in radioassays were repeated using non-radioactive UDP-Glc under the following conditions: acceptor 200 µM; UDP-Glc 1 mM; UGT 2 µg; pH 7.5 (see section 2.2.14). Reactions were carried out at 30ºC overnight. UGT73C10 was used as a positive control here for glucosylation of triterpenoids (Augustin et al. 2012). Under the experimental conditions used in this assay, UGT73C10 catalysed the formation of β-amyrin glucoside, as expected (Fig. 6.8, lane 2). None of the three UGTs (AsGT14h20, AsGT24i2 and AsGT02436) for which a weak activity was measured in the presence of β-amyrin-3-*O*-Ara (table 6.1 compound 8, Fig. 6.6) accumulate a detectable product under the conditions used (Fig. 6.8, lanes 6, 8 and 9). These results may indicate that the radioassay results were likely to be attributable to background it is also possible that glucosylation of β-amyrin-3-*O*-Ara did not occur in conditions used in the present assay or TLC analysis may be not sensitive enough to detect product formation. Further investigations on optimal enzymatic conditions in addition to repetition of radioassay analysis are required here to confirm or refute activities of these enzymes toward of β-amyrin-3-*O*-Ara. Surprisingly, AsGT24i2 and AsGT05827 both catalysed the formation of products using capsidiol as acceptor, despite the weak signal displayed in the radioassay (Fig. 6.8, lanes 13 and 14); AsGT24i2 shows a small conversion of capsidiol and no activity was detected for AsGT05827 (Fig. 6.6). As suggested by the radioassay, AsGT05827 was also able to form a product with α-bisabolol (Fig. 6.8, lane 18). α-(-)-Bisabolol was not visualised on the TLC (Fig. 6.8, lanes 16-18), and appears to have been lost over processing of the samples. Repetition of the reaction confirms the formation of a new product more polar than bisabolol, consistent with bisabolol glucoconjugate (Fig. 6.9). Once again, bisabolol was not detected in control and reaction samples (Fig. 6.9, lanes 2-3). It is unclear why bisabolol is lost through the present experimental process; α-bisabolol boiling point is 153°C (at 10mm Hg) and it is unlikely that this substrate evaporates under the reaction conditions used.
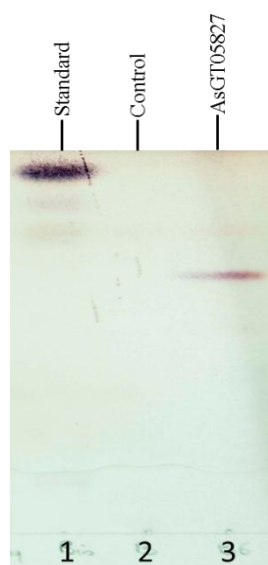
Figure 6.9: Activity of AsGT05827 towards α-(-)-bisabolol. Glucosylation reaction were performed at 35°C for 12 hours using 1mM of UDP-Glc, 200 μM of bisabolol and 2 μg of recombinant AsGT05827. Glucosylated products were extracted in ethyl acetate. Lane 1, bisabolol standard; lane 2; reaction control without enzyme; lane 3, bisabolol glucosylation reaction catalysed by AsGT0582.

TLC analysis using non-radioactive UDP-Glc was also carried out with different potential flavonoid acceptors. These assays confirmed the activity of AsGT24i2 towards the three flavonoids used in this study. Complete conversion of kaempferol and quercetin was observed with AsGT24i2 (Fig. 6.8, lanes 4 and 11); tricin was partially converted under the same reaction conditions (Fig. 6.8, lane 15). AsGT02436 was able to catalyse substantial conversion of kaempferol to a product, while tricin was converted to a lesser extent (Fig. 6.6, lanes 5 and 16). Under the conditions tested, AsGT05827 converted only a small fraction of kaempferol and no activity was detected towards tricin (Fig. 6.8, lanes 6 and 17). Minor conversion of kaempferol was also observed for the homologous enzymes UGT23586A and UGT23586B (Fig. 6.8, lanes 7, 8).

These results generally confirmed activities measured by the radioassay experiments except for AsGT05827. Radioassays suggested AsGT05827 was the most active enzyme toward tricin, and activity towards kaempferol was similar to AsGT24i2 and AsGT02436 (Fig. 6.6, table 6.1); on the contrary, no product was detected by TLC analysis when using tricin as an acceptor for AsGT05827 and only traces of product were detected for kaempferol. These results tend to show that radioassays of AsGT05827 are misleading; this is certainly related to the signal obtained in the absence of acceptor (Fig. 6.6).

Identification of the products was conducted via LC-MS analysis (Fig. 6.10; method described in section 2.2.25). Further evidence of the accumulation of monoglucosylated flavonoids catalysed by AsGT24i2, AsGT02436 and AsGT23586B was obtained. In the case of the kaempferol glucosides, a predicted parental ion was detected (m/z 449) with a major fragment (m/z 287) corresponding to kaempferol produced by a loss of a glucose residue (162 Da) (Fig. 6.10.B). A similar fragmentation pattern was observed for tricin glucoside, the major fragment (m/z 331) corresponds to a glucose loss of the parental ion (m/z 493). For all of the samples analysed, more than one peak corresponding to monoglucosylated products was detected, suggesting the formation of several isomers (Fig. 6.10. B). In the case of kaempferol, multiple products were formed by AsGT23586B with a major peak (Rt 5.39 min) and several other products (Rt 5.58 min, Rt 5.28 min, Rt 5.24 min) plus a few other minor peaks. AsGT24i2 and AsGT02436 reaction products gave simpler and similar patterns with one major peak (Rt 5.36 min and Rt 5.44 min respectively) and a smaller peak eluting slightly later (Rt 5.25 min and Rt 5.30 min respectively) plus a few minor products. In the case of tricin, AsGT02436 catalyses the formation of three monoglucosylated isomers (Rt 5.53 min > Rt 5.45 min ≈ Rt 5.39 min). These results may reflect a higher regioselectivity of AsGT24i2 and AsGT02436 compared to AsGT23586B. Mass spectrometry analysis of sesquiterpenoid products remain to be done, initial attempt results in absence of signal as described for triterpenoid glycoside (see section 4.2.4).

Taken together, these data indicate that several of the *A. strigosa* UGTs within the collection are able to catalyse the formation of multiple isomers of monoglucosylated flavonoids; glucosylation of sesquiterpenoids were also observed. Full structural determination of the products generated by these enzymes will require improved methods for separation of the glucosylated products, coupled with the use of flavonoid glucoside standards and proton NMR analysis. The identification of the glycosylated sesquiterpenoids generated in these experiments remains to be performed. TLC analysis did not support the glucoconjugation of arabinosylated β-amyrin suggested by radioassay for AsGT24i2 and AsGT02436 (21% and 9% of their activity towards TCP respectively). None of the enzymes included in these experiments were able to glucosylate lupeol or β-amyrin under the reaction conditions used.

Figure 6.10: LC-MS analysis of reaction products with *A. strigosa* UGTs and flavonoid acceptors. A. UV spectrum of reaction products of AsGT02436 with kaempferol; detection at 380 nm. Several monoglucosylated products (Rt 5.17 min and 5.33 min) are formed; unreacted kaempferol (Rt 7.15 min) is detected. **B.** Fragmentation of the parental ion (m/z 449) released kaempferol (m/z 287) after the loss of a hexose residue (m/z 162). **C.** Zoom in total ion chromatogram of monoglucosylated products of flavonoids reveals multiple peaks.

6.3   Conclusion

In this chapter the evaluation of a rapid and simple method for detecting UGT activity was reported. The procedure was developed initially using the generic acceptor TCP and proved to be reproducible. The radioassay was then applied to assess the activity of *A. strigosa* UGTs towards terpenoids and flavonoids. Enzymatic activities detected in the radioassay were generally supported by enzymatic assay with cold substrates and mass spectrometry analysis. This preliminary study suggests *A. strigosa* UGTs catalyse the formation of sesquiterpenoid glucosides, activities which have not been reported previously. The highest activities were reported towards flavonoids and TCP and indicate differential promiscuity and regioselectivity of glucosyl transfers catalysed by *A. strigosa* enzymes.

6.3.1   A radioassay for rapid glycosyltransferase acceptor screening

The radioassay used in this study to assess the activities of UGTs towards potential acceptors presents some advantages and some disadvantages. The protocol as optimised and applied to recombinant *A. strigosa* UGTs is rapid and requires only simple laboratory equipment (benchtop centrifuge, ion exchange resin, tubes and pipettes, and a scintillation counter) and handling. Importantly, it produces only scintillation waste and a small amount of solid radioactive waste, making it easy to handle when it comes to waste disposal. The protocol used here involves detection of the radioactivity in solution after the binding of the unreacted sugar donor to the resin. Consequently product formation is not directly monitored and this is the source of the major drawbacks of this radioassay. This analysis requires alternative enzymatic assays to support product formation. Also, the detection of radioactivity in the absence of an acceptor for some enzyme preparations (e.g. AsGT21p16, AsGT27f7 and AsGT05827) may be due to glucoconjugation of a component of the reaction mixture (e.g. imidazole or ethanol traces) or hydrolysis of the sugar donor by the enzyme itself (misfolded glycosyltransferases might be responsible for hydrolytic activity). Additionally, the use of ion exchange resin limits the application of this approach to only neutral or cationic acceptors, ruling out the use of acidic acceptors.

The radioassay strategy proved to be effective for some UGT/substrate combinations where there was a high level of activity towards the substrate in question, i.e. the activity of AsGT24i2, AsGT02436 and AsGT05827 towards flavonoids. The sensitivity of the radioassay was not sufficient to detect lower activities shown by AsGT23586A and AsGT23586B towards flavonoids, certainly due to the background noise remaining after optimisation. Weaker activities were detected only with TLC analysis of enzymatic reactions with cold substrates assays. This include AsGT23586A and AsGT23586B activities over quercetin. In case of weaker activities traditional analysis of radioassays with TLC separation and detection by phosphorimager may be more informative (Augustin et al. 2012; Caputi et al. 2008).

6.3.2 Flavonoid glycosyltransferase activities of *A. strigosa* UGTs and regioselectivity

The highest conversions of flavonoid substrates under the condition tested were observed for *A. strigosa* UGTs of family UGT73, AsGT24i2 and AsGT02436. Flavonoid glycosyltransferases belonging to the UGT73 constitute the clade IIIa of flavonoid glycosyltransferases (Fig. 3.1) (Noguchi et al. 2009). They are generally regarded as flavonoids-7-*O*-glycosyltransferases, catalysing regiospecific glycosylation at the C-7 position of flavonoids (Yonekura-Sakakibara and Hanada 2011). This is supported by several studies reported the characterisation of flavonoid-7-*O*-glycosyltransferases belonging to the UGT73 family (Fig. 6.11) (Hirotani et al. 2000; Kim et al. 2006).
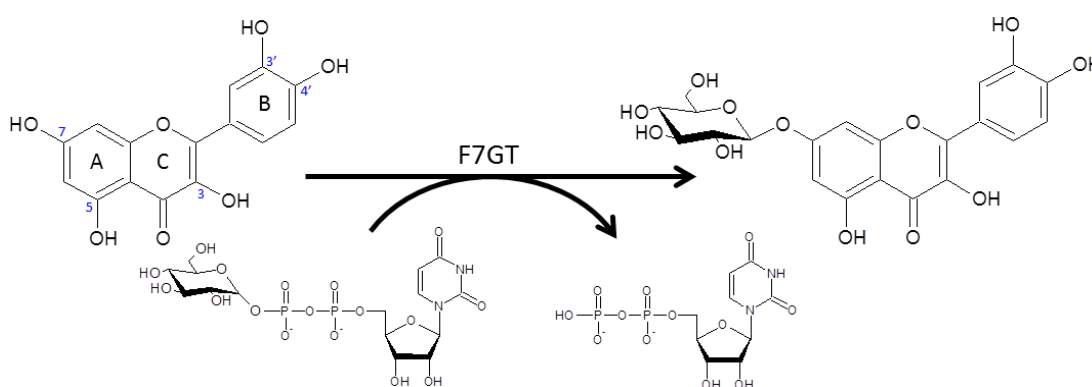
Figure 6.11: Activity of flavonoid 7-*O*-glucosyltransferase over quercetin.

Closer examination of the functional studies performed with UGT73s suggests the regiospecificity over flavonoids is not strictly conserved throughout the whole UGT73 family. Lim et al. (2004) performed glucoconjugation of quercetin with a collection of *A. thaliana* UGTs and only two out of the seven UGT73s tested displayed a preference for the C-7 position of quercetin. UGT73B4 and UGT73B5 show the highest activity (12.14 and 11.58 nkat/mg protein respectively) toward quercetin with regiospecificity for the C-3 position of the acceptor. The only functional analysis that has been carried out so far for a monocot UGT73 was on a glycosyltransferase from rice (RUGT-5). The RUGT-5 recombinant enzyme was able to form multiple monoglucosylated products over flavonoids. Proton NMR analysis suggested that RUGT-5 is able to add glucose onto positions C-3, C-7 or C-4' of flavonoids (Ko et al. 2006). More functional analyses are required to fully understand the regioselectivity of UGT73s over flavonoids acceptors and evolution of this mechanism.

The large collection of *A. strigosa* recombinant UGT73s generated in this thesis is a valuable basis to further understand the regiospecific glycosylation of flavonoids. Preliminary biochemical studies performed with *A. strigosa* AsGT14h20, AsGT16f23, AsGT24i2 and AsGT02436 suggest that their aptitude to glucosylate flavonoids is conserved at least in two of the five phylogenetic clusters identified previously (e.g clusters M3 and M5; see section 3.2.4). Authentic kaempferol glucosides standards are required to help determine regiospecificity of these enzymes, those compounds are commercially available. Structure determination will be confirmed by high field NMR and MS-MS approaches.

It is unclear whether the major products of AsGT24i2 and AsGT02436 are identical or correspond to independent structural isomers: the similarity of LC-MS peak patterns might reflect similar regiospecificities of these enzymes. Both of these enzymes show activity over tricin, the absence of an hydroxyl group in position C-3 of tricin may suggest AsGT24i2 and AsGT02436 are not showing preference for C-3 position of flavonoids. In contradiction, RUGT-5 is part of the cluster M3 sharing 72% identity with AsGT24i2 (Fig. 3.3); RUGT-5 generates kaempferol-3-*O*-glucoside as a major product when incubated with kaempferol (Ko et al. 2006). Interestingly, RUGT-5 has a quercetin-4'-*O*-glucosyltransferase activity and uses the C-3 position only in the absence of 3' hydroxyl group (e.g. kaempferol) (Fig. 6.12).

It shows that hydroxyl position on flavonoids influence the sugar specificity, therefore glycosylated positions on tricin and kaempferol may be different with *A. strigosa* UGT73s. Radioassays suggest AsGT02436 is not active with quercetin; its activity may be affected by the presence of the hydroxyl group in position C-3' of flavonols. On the contrary, AsGT24i2 uses quercetin as an acceptor and appears to convert the entire pool of acceptor (kaempferol or quercetin) under the conditions used (e.g. cold assays).
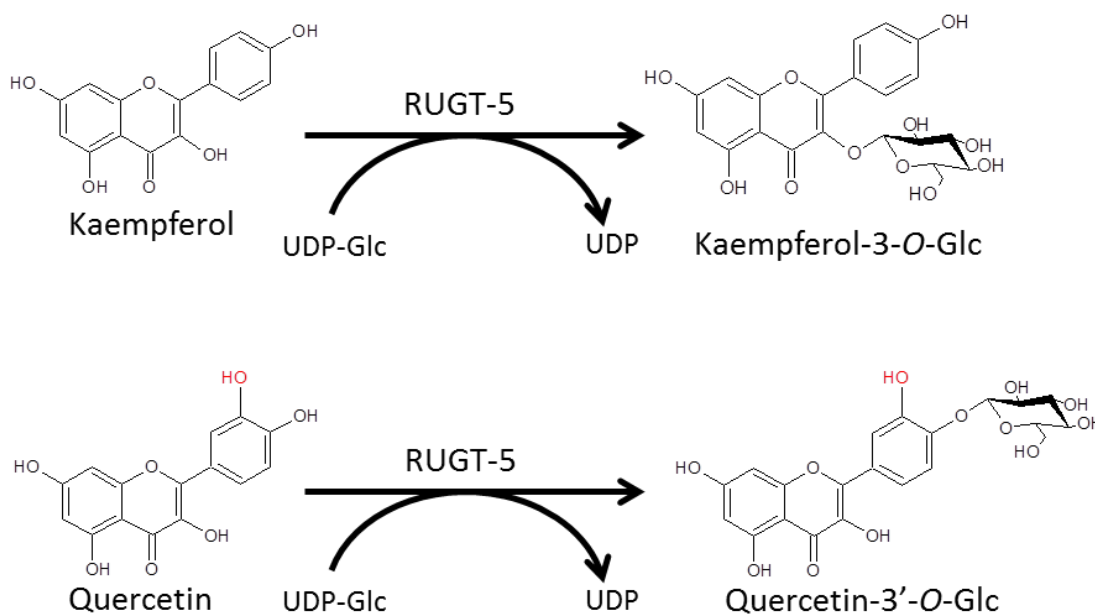
Figure 6.12: The regioselectivity of RUGT-5 is affected by the presence of 4' hydroxyl group (Ko et al. 2006).

Quercetin is known to have multiple beneficial effects on human health, including anti-inflammatory, anti-oxidative, anti-mutagenic and neuroprotective activities (Leiherer et al. 2013; Ossola et al. 2009). Glucosylation has been shown to enhance absorption of quercetin by humans (Hollman et al. 1999). Also the nature and position of glycosylation are important factors for their absorption (Heim et al. 2002), making regioselective biocatalysts from cereals biotechnological tools of great interest to modify dietary flavonoids profiles. AsGT14h20 and AsGT16f23 are unable to glycosylate the flavonoids tested under the conditions used. Taken together, these results may suggest functional divergence between the monocot phylogenetic branches defined in section 3.2.4. More flavonoid acceptors need to be tested in order to answer this question.

AsGT23586A and AsGT23586B convert kaempferol to a limited extent under the conditions used. These enzymes are part of UGT88 family, included in clade IIIb of the flavonoid glycosyltransferases. UGT88s are also regarded as flavonoids-7-*O*-glycosyltransferases (Frydman et al. 2013; Noguchi et al. 2009). LC-MS results suggest that AsGT23586B produces multiple isomers of kaempferol monoglucoside. Further work had to be done to separate those isomers and identified the products to decipher *A. strigosa* UGTs regiospecificity.

6.3.3 Sesquiterpene glycosyltransferase activities - implications and perspectives

Among the *A. thaliana* UGTs that are active on sequiterpenoids, ester-forming enzymes from families UGT74, UGT75 and UGT84 are able to catalyse the formation of glucoconjugates using the acidic sesquiterpenes artemisic acid and retinoic acid as substrates. Farnesol is glucosylated by enzymes from families UGT71, UGT73, UGT85 and UGT88 (Caputi et al. 2008). Here we have shown that AsGT05827 from the newly characterised group O (Caputi et al. 2011) is able to catalyse glucosylation of both non-acidic sesquiterpenes used in this study (Fig. 6.7, lanes 14, 18). AsGT05827 shows promiscuity, with activity towards structurally dissimilar α-bisabolol and capsidiol, which are monocyclic and dicyclic sesquiterpenes, respectively. A single site for *O*-glucosylation is available for α-bisabolol, therefore it is quite likely that the product formed is α-bisabolol 7-*O*-glucoside (Fig. 6.13). Nonetheless, a proper elucidation of the product structure is needed to exclude C-glucosylation or formation of a diglucosidic chain. However, since AsGT05827 has been shown to have TCP *O*-glucosyltransferase activity (section 4.2.3.3), we can speculate that sesquiterpenes *O*-glucosides are likely to be formed in these assays.
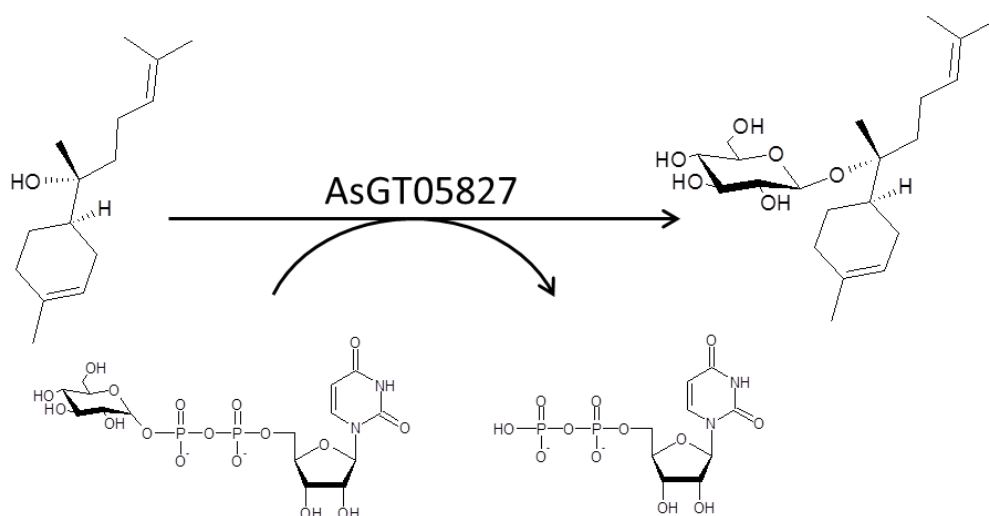
Figure 6.13: Suggested activity of AsGT05827 with bisabolol as acceptor.

α-Bisabolol has a cytotoxic effect on glioma cells, a cancer cell type responsible for 80% of brain tumours (Cavalieri et al. 2004). Monoglycosylated forms of α-bisabolol have been prepared by chemical synthesis and tested against various cancerous cell lines (Piochon et al. 2009). In most cases, glycoconjugates of α-bisabolol exhibit a higher cytotoxicity than the aglycone. α-Bisabolol-*O*-β-D-glucoside shows higher cytotoxicity against several cell lines; LD_{50} was up to 36% smaller (LD_{50} 130 μM of α-bisabolol and 96 μM of α-bisabolol-Glc over human primary glioblastoma cell line U-87). In this study, α-bisabolol rhamoside was the most active compound overall (LD_{50} 46 μM of α-bisabolol-Rha) (Piochon et al. 2009). Additionally, the modification of the physicochemical parameters of α-bisabolol induced by glycosylation are likely to improve the capacity of the compound to cross the blood-brain barrier and access the cancer cells, this is supported by *in silico* predictions (Piochon et al. 2009). Consequently, AsGT05827 might be an excellent candidate for a synthetic biology approach to produce α-bisabolol *O*-β-D-glucoside. A kinetic analysis is required to determine the affinity and activity of the AsGT05827 towards α-bisabolol.

6.3.4   General conclusion

Screening of a subset of potential acceptors with the collection of *A. strigosa* recombinant UGTs led to the discovery of new catalytic activities with potential biotechnological applications. These preliminary biochemical studies, together with phylogenetic analysis, suggest a complex evolution of substrate recognition and regioselectivity displayed by monocot UGTs. The biological significance of the observed activities remains to be addressed. Heterologous expression of AsGT24i2 and AsGT02436 in *A. thaliana* or *N. benthamiana* and analysis of flavonoid profiles might give clues about the potential involvement in flavonoids biosynthesis of these enzymes. Their respective expression profiles suggest divergent functions *in planta* AsGT02436 is expressed only in roots while AsGT24i2 is ubiquitous. The nature of the flavonoid content present in *A. strigosa* tissues expressing those enzymes is also essential in elucidating the *in vivo* functions of this class of *A. strigosa* UGTs.

Chapter 7 – General conclusion and future work

Plants family one UDP-dependant glycosyltransferases (UGTs) are responsible for the glycosylation of a multitude of endogenous or exogenous small molecules. Depending on the nature of their acceptors, UGTs may be involved in cellular homeostasis, synthesis of bioactive compounds or detoxification mechanisms. This thesis describes the identification and functional investigation of UGTs expressed in oat root tips.

7.1 Identification and sequence analysis of UGTs expressed in oat root tips

In chapter 3 a comprehensive collection of UGTs that are expressed in oat root tips was unveiled after mining of a transcriptomic dataset. A total of 110 unique UGT-like transcripts were identified by tBLASTn searches, 53 of which correspond to full-length sequences. The number of UGT-like sequences detected in oat root tips is in accordance with recent genome-wide analyses of plant UGTs in other species (Caputi et al. 2011; Yonekura-Sakakibara and Hanada 2011).

Further analyses were carried out on UGT transcripts that had been assembled as full-length sequences. A phylogenetic tree was constructed with a collection of functionally characterised UGTs from other plant species in which the oat full-length UGT amino acid sequences were included. The oat UGTs were heterogeneously distributed across the UGT phylogeny, the occurrence of clusters of oat UGT sequences illustrating the lineage-specific expansion of UGTs as has been observed in other plant species (Caputi et al. 2011). The number of UGTs contained in each of the 16 phylogenetic groups reported by Ross et al. (2001) and Caputi et al. (2011) differs across species as a result of the lineage-specific expansion of UGTs. In oat, groups L, E and D were particularly overrepresented in root tips, suggesting a specific requirement for functional diversification of the corresponding enzymes (those groups are particularly involved in glycoconjugation of triterpenoids/flavonoids [D], flavonoids/monolignols [E] and carboxylic acids [L]). In contrast, no sequences from groups B, F and M were discovered in the oat root tip transcriptome. This is suggesting these enzymes are not involved in root physiology

in laboratory conditions or that a gene loss event lead to the disappearance of these groups in oat.

Avenacin A-1 is a triterpenoid saponin with antifungal properties that accumulates in oat root tips. Characterisation of the complete biosynthetic pathway of avenacin A-1 including the steps required for glycosylation will enable engineering of the anti-fungal properties attributable to avenacin into susceptible crop species. The enzymatic processes required for the synthesis of the avenacin trisaccharide are not unknown. The majority of UGTs that are active towards triterpenoids or related compounds (e.g. phytosterols, glycoalkaloids, brassinosteroids) belong to phylogenetic group D (equivalent to family UGT73). In order to gain knowledge of potential avenacin glycosyltransferase candidates a phylogenetic reconstruction of group D was performed with entire UGT73 families from rice and thale cress enriched with functionally characterised enzymes from other plant species and containing all oat UGT73s discovered in this work. The tree topology indicates divergent evolution of UGT73s in monocots and dicots; dicot UGT73s are split into two monophyletic branches, while monocot UGT73s are spread over several branches. Considering this and the low number of functional studies performed on monocot species it is not clear whether the monocots and dicots from group D have evolved similar functions. Additional functional studies of monocot UGT73s are required to understand the evolution of this phylogenetic group across lineages. The functional analysis in chapters 4 and 5 suggests that triterpenoids and flavonoids are glycosylated by oat UGT73s, supporting the idea that similar functions are performed by monocot and dicot UGTs from group D. AsGT16f23 was part of a cluster (M1) orthologous to the dicot cluster containing triterpenoid-3-*O*-glucosyltransferases from *B. vulgaris* (D1), making it a candidate for avenacin glycosylation. However, functional analysis suggests that AsGT16f23 is not active towards triterpenoids. A subset of 17 UGTs - candidates for avenacin glycosylation – was selected based on phylogenetic analysis complemented with gene expression profile and proteomic analyses. The differential distribution of UGTs in oat root tips and elongation zones and the expression of *Ugt* genes were compared to characterised avenacin biosynthetic genes. Which are known to be tightly co-expressed in epidermal cells of oat root tip. From the 17 UGTs selected 13

were cloned from oat root tip cDNA and inserted into expression vectors for functional analysis.

7.2   <u>Functional investigation of *A. strigosa* UGTs - results and perspectives</u>

Plant UGTs are generally very selective towards their sugar donors, and their activity dramatically decreases in the presence of non-natural UDP-sugar substrates. Consequently, the understanding of sugar donor specificity is an essential prerequisite for the functional analysis of UGTs. In chapter 4, a method was designed to investigate sugar specificity for UGTs of unknown functions. This method takes advantage of the promiscuity shown by plant UGTs towards their acceptor substrates and the broad proportion of plant UGTs able to use TCP as an acceptor. TCP-β-D-glucopyranose and TCP-α-L-arabinopyranose were respectively produced by the *N*-methylanthranilate-*O*-glucosyltransferase (SAD10) and the flavonol-3-*O*-arabinosyltransferase (UGT78D3). This procedure was applied to the oat UGTs; all of the recombinant glycosyltransferases that were active towards TCP preferentially used UDP-Glc as sugar donor. Nevertheless, TCP arabinosyltransferase activity was detected for 9 of the 15 enzymes tested; among them AsGT14h20, AsGT14h21, AsGT24i2 and AsGT25n16 possess a substantial secondary activity with UDP-Ara compare to UDP-Glc. The present analysis may be completed using more sugar donors, including UDP-Rha, UDP-Man, UDP-GlcA, commonly used by plant UGTs (Bowles et al. 2005; Osmani et al. 2009). Based on this procedure, a competition assay will also be interesting to develop using a mix of UDP-sugars as substrates. Catalytic behaviour of plant UGTs in the presence of multiple sugar donors has not been investigated so far and may shed new light on sugar selectivity of UGTs *in planta*.

In chapter 5, a radioassay was developed for the rapid screening of glycosyltransferase acceptors. TCP was used as an acceptor for the initial development of the method, confirming the results obtained in section 4.2.3. A preliminary functional analysis was then conducted onto a subset of eight *A. strigosa* glycosyltransferases selected based on their TCP glucosyltransferase activity; flavonoids, sesquiterpenoids and triterpenoids were used as potential acceptors. Substantial activities were detected -particularly towards flavonoids - with three of

these recombinant enzymes. Most of the activities detected using the radioassay were supported by the detection of reaction products using TLC analysis, and generation of monoglucosylated flavonoids was confirmed by mass spectrometry. These preliminary results suggest a broad promiscuity of AsGT24i2 toward its acceptors; AsGT24i2 glucosylates quercetin, kaempferol, tricin and TCP to a similar extent under the conditions used and traces of glucosylated capsidiol were also detected. Higher substrate specificity was observed for AsGT02436 using kaempferol as preferential acceptor despite activities towards TCP, tricin and hederagenin. As described in section 6.2.1 radioassays performed with AsGT05827 were misleading; nevertheless, product formation was detected by TLC towards kaempferol and sesquiterpenoid acceptors. Kinetic analysis remains to be done to assess the specificity of these enzymes towards their acceptors. Kinetic analysis combined with metabolic profiling of *N. benthamiana* leaves expressing these enzymes may prove to be a powerful approach to investigate the biological significance of the activities detected in this study.

The α-bisabolol glucosyltransferase activity displayed by AsGT05827 is of great interest in the scope of developing enhanced drugs for brain cancer treatment (Cavalieri et al. 2004; Darra et al. 2008). Full identification of the reaction product as well as kinetic analysis remains to be done in order to estimate the true potential of this enzyme. Nevertheless, AsGT05827 may be an excellent template for the development of synthetic biology approaches for the generation of bisabolol glycoconjugates. Protein engineering may be used to alter sugar specificity of AsGT05827 (Hansen et al. 2009; Kubo et al. 2004; Noguchi et al. 2009; Osmani et al. 2008) considering that α-bisabolol rhamnoside has the most cytotoxic activities against cancer cell lines (Piochon et al. 2009).

7.3 *N. benthamiana* as a platform for the production of synthetic saponins

The results presented in chapter 5 confirm the accumulation of 12,13-epoxy-16-hydroxy-β-amyrin-3-*O*-β-D-glucopyranoside in *N. benthamiana* leaves that have been co-infiltrated with SAD1, SAD2 and UGT73C10 expression vectors. This is the first report of the production of a simple triterpenoid saponin in a heterologous expression system via co-expression of an OSC, a P450 and an UGT, three multi-gene families that all have central roles in triterpenoid biosynthesis (Augustin et al. 2011; Sawai and Saito 2011). This work is therefore a proof-of-principle that transient co-expression in *N. benthamiana* using the pEAQ-*HT* system is a valuable platform for the production of oleanane-based triterpenoid saponins. The results presented in chapter 5 together with the work of Geisler et al. (2013) also suggest the absence of endogenous activities towards the various triterpenoid compounds produced: β-amyrin, 12,13-epoxy-16-hydroxy-β-amyrin, β-amyrin-3-*O*-β-D-glucopyranoside and the likely 12,13-epoxy-16-hydroxy-β-amyrin-3-*O*-β-D-glucopyranoside. Activity of endogenous enzymes remains a major issue in synthetic biology approaches (Moses et al. 2013; van Herpen et al. 2010); the absence of endogenous modification of the triterpenoids produced so far constitute an excellent basis for the development of a saponin production platform in *N. benthamiana*.

Development of the present platform may be extremely rewarding considering the various industrial applications of saponins (e.g. food sweeteners, surfactants, adjuvents, antimicrobial agents). Triterpenoid-modifying enzymes reported in the literature (see chapter 1) may be used for combinatorial biosynthesis of complex and diverse new-to-nature saponin structures (Moses et al. 2013). The productivity of the system remains to be precisely evaluated but the initial production of 12,13-epoxy-16-hydroxy-β-amyrin was more than 1 mg/g dry weight (Geisler et al. 2013). Higher yields may be obtained considering the "detoxification" effect of glucosylation on the triterpenoid products of SAD1 alone or SAD1 plus SAD2 suggested by the phenotypes observed in leaves. Metabolic engineering of the MVA and MEP pathway has been used successfully to enhance production of carotenoids in tomato (Enfissi et al. 2005); a similar approach might be developed to increase the flux of triterpenoid precursors.

7.4 Implications of the present work for avenacin glycosylation

One of the initial aims of this work was to increase our understanding of avenacin trisaccharide formation. To do so avenacin glycosyltransferase candidates were identified based on phylogenetic, root proteomic and gene expression analysis in chapter 3. Triterpenoid-3-*O*-arabinosylation or triterpenoid-3-*O*-arabinoside glucosylation activities were not detected by TLC analysis of *in vitro* assays using multiple experimental conditions in chapter 4. Weak radioactive signals were detected in solution when incubating β-amyrin-3-*O*-α-arabinopyranoside with radiolabelled UDP-Glc in chapter 5; it is unclear whether the radioactive signal is due to enzymatic activity or background noise in these assays. Potential glucosylated products remain to be detected and identified in order to establish if these enzymes (AsGT24i2 and AsGT02436) may be involved in glucosylation of avenacin; this will be achieved by chromatographic separation of the reaction products from the radioassays and by structural analysis (H^1 NMR, mass spectrometry).

Identification of triterpenoid glycosyltransferases involved in saponin biosynthesis is challenging due to several factors: the unavailability of suitable acceptors, the hydrophobic nature of triterpenes and the problems of detection of the potential products encountered with mass spectrometry. Obtaining the hypothetical natural acceptor of the glycosyltransferases (deglycosylated desacyl avenacin) will improve our chances of discovering avenacin glycosyltransferases. This compound can be obtained by hydrolysis of the trisaccharide of desacyl avenacin accumulated by *sad7* mutant (Mugford et al. 2009). Obtaining such a compound is not trivial. The avenacin epoxide is labile under acidic conditions and 0.1% trifluoroacetic acid has been used to form carbonyl-avenacin after conversion of the epoxide (Geisler et al. 2013); consequently acidic hydrolysis is a poor option. Therefore enzymatic hydrolysis might be regarded as the best option; the specific glycosidase avenacinase may use desacyl avenacin as a substrate (Osbourn et al. 1995). The promiscuous β-glucosidase from almonds is another alternative and may conceivably release α-L-arabinose as well as β-D-glucose (Gachon et al. 2004).

The absence of an enzyme showing a preference for UDP-Ara over UDP-Glc raises the question of the presence of avenacin arabinosyltransferase in our collection of recombinant enzymes. The absence of specific arabinosyltransferase activity may indicate that the candidate screening for avenacin glycosyltransferase performed in chapter 3 failed to identify avenacin arabinosyltransferase or that this enzyme was part of the subset of potential candidates that were not cloned in section 4.2.1. Alternatively, the hypothetical model of trisaccharide formation could be incorrect. A peculiar mechanism of glycosylation may occur in avenacin mechanism. Recently a vacuolar transglucosidase has been identified from rice (Luang et al. 2013). This enzyme transfer sugars from phenolic acid esters (leaving group pK_a 4.6-5.8) to a variety of acceptors (flavonoids and phytohormones). Such a mechanism could occur in oat vacuoles where the presence of the acyl glucose donors benzoyl-β-D-glucose and N-methyl-anthranyloyl-β-D-glucose (leaving group pKas 4.2 and 2.1, respectively) is known to be required for SAD7 activity. Other alternative routes are also possible including blockwise addition of activated preformed trisaccharide. In order to confirm the initial hypothetical pathway, activity assays may be performed on protein crude extract from oat root tips to know if any arabinosyltransferase activity towards suitable triterpenoids acceptor can be detected.

The recent gain of knowledge in glycosyltransferases involved in glycosyl transfer to a sugar moiety of a glycoside (GGTs) is of great interest for identification of avenacin glucosyltransferases and for understanding the synthesis of natural product oligosaccharide chains more widely. Indeed, only a few GGTs were characterised from the family UGT94 at the time that I started my PhD (Masada et al. 2009; Noguchi et al. 2008; Osmani et al. 2008). Over the last three years many new GGTs have been discovered in group A, not only in the UGT94 family (Itkin et al. 2013; Nagatoshi et al. 2012; Ono et al. 2010b) but also in the UGT79 family (Frydman et al. 2013; Yonekura-Sakakibara et al. 2012) and UGT91 family (Shibuya et al. 2010), suggesting that GGT activity appeared before the split between UGT79, UGT91 and UG94 (Yonekura-Sakakibara et al. 2012). These families are therefore obvious targets for discovery of glucosyltransferases involved in triterpene glycoside biosynthesis. The expression profile analysis of AsGT01332 and AsGT18279 – belonging to the group A (Fig. 3.1) – will bring new light on the potential role of these enzymes in avenacin biosynthesis. Of note, AsGT01332 has been detected in

the proteomic analysis and data suggests this protein is more abundant in root tips (Fig.3.5).

The identification of enzymes involved in avenacin trisaccharide formation is of importance in the scope of engineering take-all resistant crops. Also, antimicrobial activity is common for saponins and alternative glycosylation pattern is also of interest to improve resistance of susceptible crop species. An alternative approach to the enzymatic synthesis of avenacin trisaccharide may be the engineering of glycosyltransferases with required catalytic properties. In the past, domain swapping approaches have taken advantage of the conserved three dimensional structures of UGTs to generate new activities (Cartwright et al. 2008; Hansen et al. 2009). A combination of approaches – point mutation of crucial residues for sugar donor recognition, domain swapping, PSPG motif exchange/modification – may be applicable to UGT73C10 (triterpene-3-*O*-glucosyltransferase from *B. vulgaris*) to modify its sugar specificity.

7.5   Concluding remarks

The present work suggests that functional evolution of UGTs in monocots and dicots has taken rather different routes reflected by discrete expansion of UGT families. Further functional characterisation of monocot UGTs is required in order to understand the primary structure-activity relationships of monocot UGTs. Functional investigation of *A. strigosa* UGTs has uncovered catalytic activities of potential biotechnological interest. The development of a powerful *N. benthamiana*-based production platform for synthetic saponins offers great perspectives for the generation of improved/new-to-nature compounds with valuable properties.

Supplementary data

UGTs	GenBank accession	Group
UGT71B1	AB025634	E
UGT72B1	AF360262	E
UGT73B1	AL021961	D
UGT74B1	AC002396	L
UGT75B1	AV790637	L
UGT76B1	AC073395	H
UGT78D1	AC009917	F
UGT79B1	AB018115	A
UGT80A2	AC016827	n.d
UGT80B1	AC007203	n.d
UGT81A1	AL031004	n.d
UGT82A1	AP002046	N
UGT83A1	AC011664	I
UGT84A1	CP002687	L
UGT85A1	AC006551	G
UGT86A1	AC006922	K
UGT87A1	AC004165	J
UGT88A1	AP000373	E
UGT89A1	AC006085	B
UGT90A1	AC005167	C
UGT92A1	AL353013	M

Supplementary data S.1: Table of *Arabidopsis thaliana* UGTs used as query in tBLASTn searches for mining of 454-based transcriptomic resource from *A. strigosa* root tips

Name used full length candidates	Contig name	Contig length (nt)	Number of reads
AsGT1a115	00180	953	47
AsGT3i21	02980	-	-
AsGT4h2	17583	524	5
AsGT8i4	08947	1708	76
AsGT11i11	02362	1725	79
AsGT12o13	10086	799	20
AsGT14h16	00260	1678	102
AsGT14h21	13141	1672	31
AsGT15a11	01799	1660	139
AsGT16f23	01781	1709	126
AsGT16h6	09778	1614	620
AsGT18p9	01799	1660	139
AsGT20n10	00101	1684	270
AsGT21p16	00260	1678	102
AsGT24i2	26925	915	11
AsGT25n16	10190	1375	39
AsGT27a12	22121	1562	61
AsGT27f7	16327	1690	94
AsGT28b19	10762	840	20
AsGT260	00260	1678	102
AsGT678	00678	1945	85
AsGT733	00733	1562	76
AsGT1092	01092	1594	116
AsGT1332	01332	1581	95
AsGT1670	01670	1611	61
AsGT1989	01989	1651	575
AsGT2436	02436	1552	85
AsGT3158	03158	1785	67
AsGT3999	03999	2338	81
AsGT5827	05827	1577	65
AsGT6492	06492	1619	44
AsGT6751	06751	1693	77
AsGT8947	08947	1708	76
AsGT10326	10326	1791	91
AsGT16525	16525	1838	117
AsGT17576	17576	1785	136

AsGT18279	18279	1673	98
AsGT18535	18535	1616	55
AsGT23586A/B	23586	2152	205
AsGT23781	23781	1814	44
AsGT23818	23818	1629	147
AsGT24525	24525	2297	840
AsGT26962	26962	1548	116
	00030	474	12
	00243	1136	16
	00931	923	32
	01194	1280	22
	01315	704	12
	01341	1363	19
	01461	451	4
	01577	890	45
	01599	795	8
	02132	712	16
	02699	588	9
	03295	644	10
	03883	780	10
	04347	790	15
	04598	737	7
	05602	1073	14
	05740	1712	38
	06218	690	9
	07600	1667	71
	07784	1426	24
	07903	1625	50
	08700	937	19
	10086	799	20
	10188	523	7
	10189	552	7
	10433	778	7
	10763	885	20
	10772	1466	87
	10811	349	56
	11099	822	14
	11140	496	3
	11637	1360	82
	12842	745	166

13541	549	23
14612	787	16
15275	509	4
15351	1186	41
16496	1318	33
17328	556	6
17424	843	14
17594	264	4
17673	692	18
17930	891	10
18035	966	12
18257	1541	59
18280	637	91
21401	1311	22
21862	1184	11
22027	648	8
22388	1435	93
22538	1554	118
23002	1061	23
23141	274	116
23340	1050	18
23453	466	7
24249	741	22
25504	573	7
26167	1494	23
26778	546	10
27009	1115	92
28651	1622	223
28947	804	21
29169	2453	414

Supplementary data S.2: Table of UGT-like retrieved from 454-based transcriptomic resource from *A. strigosa* **root tips.**

Name	Acc Numb	Plant Species	Group	UGT far	Activity	Publication
UGT79B1	BAA19127	*Arabidopsis thaliana*	A	UGT79	Flavonol glucosyltranferase	Tohge et al, 2005
UGT91A1	Q940V3	*Arabidopsis thaliana*	A	UGT91	n.d	Ross et al., 2001
UGT94B1	Q5NTH0	*Bellis perennis*	A	UGT94	Cyanidin-O-glucoside 2-O-glucuronosyltransferase	Osmani et al., 2008
CaUGT3	BAH80312	*Catharanthus roseus*	A	UGT94	Flavonoid glucoside 1,6-glucosyltransferase	Masada et al., 2009
Cm1,2RhaT1	AY048882	*Citrus maxima*	A	UGT94	Flavanone-7-O-glucose 1,2-rhamnosytransferase	Frydman et al., 2012
Cs1,6RhaT	DQ119035	*Citrus sinensis*	A	UGT91	Flavanone rutinoside 1,6-rhamnosyltransferase	Frydman et al., 2012
UGT94E5	BAM28984	*Gardenia jasminoides*	A	UGT94	iridoid-O-glucoside 6-glucosyltransferase	Nagatoshi et al., 2012
UGT91H4	BAI99585	*Glycin max*	A	UGT91	Triterpenoid-*O*-glucosyltransferase	Shibuya et al, 2010
RT	Q43716	*Petunia hybrida*	A	UGT81	Anthocyanidin-3-O rhamnosyltransferase	Brugliera et al., 1994
In3GGT	Q53UH4	*pomoea nil*	A	UGT91	Anthocyanidin-3-O-glucoside 1,2-glucosyltransferase	Morita et al., 2005
UGT94D1	BAF99027	*Sesamum indicum*	A	UGT94	Sesaminol 2'-O-glucoside-O-glucosyltransferase	Noguchi et al., 2008
UGT94F1	BAI44133	*Veronica persica*	A	UGT94	Anthocyanin-3-O-glucoside 1,2-glucosyltransferase	Ono et al., 2010
UGT89B1	NP_177529	*Arabidopsis thaliana*	B	UGT89	Benzoic acid glucosyltransferase	Lim and Bowles, 2004
UGT89C1	AAF80123	*Arabidopsis thaliana*	B	UGT89	Flavonol-7-O-rhamnosyltransferase	Yonekura-Sakakibara et al., 2007
UGT90A1	Q9ZVX4	*Arabidopsis thaliana*	C	UGT90	n.d	Ross et al., 2001
UGT90A7	ACB56926	*Hieracium pilosella*	C	UGT90	Flavonoid 7-O-glucosyltransferase	Witte et al., 2009
UGT73G1	AAP88406	*Allium cepa*	D	UGT73	Flavonoid glucosyltransferase	Kramer et al, 2003
UGT73J1	AAP88407	*Allium cepa*	D	UGT73	Flavonoid glucosyltransferase	Kramer et al., 2003
UGT73B3	AAM47999	*Arabidopsis thaliana*	D	UGT73	Flavonoid glucosyltranferase	Kim et al, 2006a
UGT73C6	AAD20151	*Arabidopsis thaliana*	D	UGT73	Flavonol 3-*O*-glucoside-7-*O*-glucosyltransferase	Jones et al., 2003
UGT73C5	AEC09299	*Arabidopsis thaliana*	D	UGT73	Brassinosteroid-23-O-glucosyltransferase	Poppenberger et al., 2005
UGT73C1	AEC09294	*Arabidopsis thaliana*	D	UGT73	Zeatin-O-glucosyltransferase	Hou et al., 2004
UGT73C10	AFN26666	*Barbarea vulgaris*	D	UGT73	Triterpenoid-3-*O*-glucosyltransferase	Augustin et al. 2012
UGT73C11	AFN26667	*Barbarea vulgaris*	D	UGT73	Triterpenoid-3-*O*-glucosyltransferase	Augustin et al. 2012
UGT73C12	AFN26668	*Barbarea vulgaris*	D	UGT73	Triterpenoid-3-*O*-glucosyltransferase	Augustin et al. 2012
UGT73C13	AFN26669	*Barbarea vulgaris*	D	UGT73	Triterpenoid-3-*O*-glucosyltransferase	Augustin et al. 2012
BvGT2	AAS94329	*Beta vulgaris*	D	UGT73	Flavonoid-7,4'-O-glucosyltransferase	Isayenkova et al., 2006
CaUGT2	BAD29722	*Catharanthus roseus*	D	UGT73	Curcumin glucosyltransferase	Kaminaga et al., 2004
Bet5OGT	CAB56231	*Cleretum bellidiforme*	D	UGT73	Betanidin-5-O-glucosyltransferase	Vogt et al., 1999

DicGT4	BAD52006	*Dianthus caryophyllus*	D	UGT73	Chalcononaringenin 2'-O-glucosyltransferase	Ogata et al., 2004
UGT73A5	CAB56231	*Dorotheanthus bellidiforn*	D	UGT73	Betanidin 5-O-glucosyltransferase	Vogt et al., 2002
ant3OGT	Q8H0F2	*Gentiana triflora*	D	UGT73	Anthocyanin 3'-O-glucosyltransferase	Fukuchi-Mizutani et al., 2003
UGT73F2	BAM29362	*Glycin max*	D	UGT73	Triterpenoid-*O*-glucosyltransferase	Sayama et al. 2012
UGT73F4	BAM29363	*Glycin max*	D	UGT73	Triterpenoid-*O*-glucosyltransferase	Sayama et al. 2012
UGT73P2	BAI99584	*Glycin max*	D	UGT73	Triterpenoid-*O*-glucosyltransferase	Shibuya et al, 2010
UGT73F1	BAC78438	*Glycyrhiza echinata*	D	UGT73	Isoflavonoid, glucosyltransferase	Nagashima et al., 2004
UGT73A10	BAG80536	*Lycium barbarum*	D	UGT73	catechin glucosyltansferase	Noguchi et al., 2008
UGT73K1	AAW56091	*Medicago trunctula*	D	UGT73	Triterpenoid-*O*-glucosyltransferase	Achnine et al. 2005
UGT73F3	ACT34898	*Medicago trunctula*	D	UGT73	Triterpenoid-28-*O*-glucosyltransferase	Naoumkina et al. 2010
Togt1	AAK28303	*Nicotiana tabacum*	D	UGT73	hydroxycoumarin glucosyltransferase	Fraissinet-Tachet et al., 1998
RUGT-5	XM_463383	*Oryza sativa*	D	UGT73	Flavonol-3,4'-O-glucosyltransferase	Ko et al., 2006
UGT73B6	AAS55083	*Rhodiola sachalinensis*	D	UGT73	tyrosol glucosyltransferase	Ma et al., 2007
UBGT	BAA83484	*Scutellaria baicalensis*	D	UGT73	Flavonoid-7-O-glucosyltransferase	Hirotani et al., 2000
GAME1	ADQ37964	*Solanum lycopersicum*	D	UGT73	Tomatidine-O-galactosyltransferase	Itkin et al., 2011
GAME2	ADQ37966	*Solanum lycopersicum*	D	UGT73	β1-Tomatidine-O-xylosyltransferase	Itkin et al., 2013
SGT1	AAB48444	*Solanum tuberosum*	D	UGT73	Solanidine/sterol alkaloid glucosyltransferase	Moehs et al., 1997
SGT2.1	ABB29873	*Solanum tuberosum*	D	UGT73	Solanidine glucosyltransferase	McCue et al., 2006
SGT3	ABB84472	*Solanum tuberosum*	D	UGT73	Solanine/chaconine rhamnosyltransferase	McCue et al., 2007
VaAOG	Q8W3P8	*Vigna angularis*	D	UGT73	Abscisic acid-O-glucosyltransferase	Xu et al., 2002
UGT88D3	ABR57234	*Antirrhinum majus*	E	UGT88	Chalcone 4'-O-glucosyltransferase	Ono et al., 2006
UGT71B6	BAB02837	*Arabidopsis thaliana*	E	UGT71	Abscisic acid glucosyltransferase	Priest et al., 2005
UGT71C1	AAC53226	*Arabidopsis thaliana*	E	UGT71	Benzoic acid glucosyltransferase	Lim and Bowles, 2002
UGT72B1	CAB80916	*Arabidopsis thaliana*	E	UGT72	Chlorinated phenol N-/ O-glucosyltransferase	Brazeir-Hicks et al, 2007
UGT88A1	AEE75831	*Arabidopsis thaliana*	E	UGT88	n.d	Ross et al., 2001
UGT72E2	AED98252	*Arabidopsis thaliana*	E	UGT72	Monolignol 4-O-glucosyltransferase	Lanot et al.,2006
UGT72D1	Q9ZU72	*Arabidopsis thaliana*	E	UGT72	n.d	Ross et al., 2001
BvGT1	AAS94330	*Barbarea vulgaris*	E	UGT71	Flavonoid-O-glucosyltransferase	Isayenkova et al., 2006
GmIF7GT	NP_001235161	*Glycine max*	E	UGT88	Isoflavone-7-O-glucosyltransferase	Noguchi et al., 2007
UGT72B11	ACB56923	*Hieracium pilosella*	E	UGT72	Flavonol-O-glucosyltransferase	Witte et al., 2009

GTase1	BAF75917	*Ipomoea nil*	E	UGT71	Phytohormones-O-glucosyltransfersase	Suzuki et al., 2007
4CGT	BAE48240	*Linaria vulgaris*	E	UGT88	Chalcone glucosyltransferase	Ono et al., 2006
UGT88A4	ABL85471	*Maclura pomifera*	E	UGT88	Coumarins glucosyltransferase	Tian et al., 2006
UGT72L1	ACC38470	*Medicago truncatula*	E	UGT72	Epicatechin 3'-O-glucoside	Pang et al., 2008
UGT71G1	AAW56092	*Medicago trunctula*	E	UGT71	triterpenoid-*O*-glucosyltransferase	Achnine et al. 2005
NtGT1b	BAB60721	*Nicotiana tabacum*	E	UGT71	Flavonoids and coumarins glucosyltransferase	Taguchi et al., 2001
UGT706C1	BAB868090	*Oryza sativa*	E	UGT88	Kaempferol and quercetin glycosyltransferase	Ko et al., 2007b
UGT706D1	BAB868093	*Oryza sativa*	E	UGT88	Flavone and flavanone glycosyltransferase	Ko et al., 2007b
UGT707A3	BAC83989	*Oryza sativa*	E	UGT71	Kaempferol and quercetin glycosyltransferase	Ko et al., 2007b
UGT88D7	BAG31948	*Perilla frutescens*	E	UGT88	Flavonoid 7-O-glucuronosyltransferase	Noguchi et al., 2009
AS	Q9AR73	*Rauvolfia serpentina*	E	UGT72	Hydroquinone glucosyltransferase (Arbutin synthase)	Hefner et al., 2002
UGT72B14	ACD87062	*Rhodiola sachalinensis*	E	UGT72	Tyrosol glucosyltransferase	Yu et al., 2011
RhGT1	BAD99560	*Rosa hybrida*	E	UGT88	Anthocyanidin 5,3-O-glucosyltransferase	Ogata et al., 2005
ScUGT5	BAJ11653	*Sinningia cardinalis*	E	UGT88	3-Deoxyanthocyanidin glucosyltransferase	Nakatsuka & Nishihara., 2010
SlUGT5	HM209439	*Solanum lycopersicum*	E	UGT72	Phenols glucosyltransferase	Louveau et al., 2011
F7GT2	BAH14962	*Torenia hybrid cultivar*	E	UGT88	flavonoid 7-O-glycosyltransferase	Noguchi et al, 2009
F7GT1	BAH14961	*Torenia hybrid cultivar*	E	UGT88	Flavonoid 7-O-glucuronosyltransferase	Noguchi et al., 2009
UGT88D8	BAH47552	*Veronica persica*	E	UGT88	Apigenin-7-O-glucuronosyltransferase	Ono et al., 2010
UGT78D1	AAF19756	*Arabidopsis thaliana*	F	UGT78	Flavonol 5-O-rhamnosyltransferase	Jones et al., 2003
UGT78D3	AED92375	*Arabidopsis thaliana*	F	UGT78	Flavonol 5-O-arabinosyltransferase	Yonekura- et al., 2003
AcGaT	BAD06514	*Aralia cordata*	F	UGT78	Anthocyanin 3-*o*-galactosyltransferase	Kubo et al. 2004
F3OGT	AAS00612	*Citrus sinensis*	F	UGT78	Flavonoid 3-O-glucosyltransferase	Lo Piero et al., 2005
an35GT	BAF49289	*Clitoria ternatea*	F	UGT78	Anthocyanin 3',5'-O-glucosyltransferase	Noda et al., 2003
Ct3GT-A	BAF49297	*Clitoria ternatea*	F	UGT78	Anthocyanin 3-O-glucosyltransferase	Noda et al., 2004
DicGT1	BAD52003	*Dianthus caryophyllus*	F	UGT78	Flavonoid-3-O-glucosyltransferase	Ogata et al., 2004
DicGT3	BAD52005	*Dianthus caryophyllus*	F	UGT78	Flavonol 3-O-glucosyltransferase	Ogata et al., 2004
F3galtase	BAI40148	*Diospyros kaki*	F	UGT78	Flavonoid 3-O-galactosyltransferase	Ikegami et al., 2009
3GalT	BAF49284	*Eustoma exaltatum*	F	UGT78	Flavonoid 3-O-galactosyltransferase	Noda et al., 2002
Ufgt	AAU09443	*Fragariax ananasas*	F	UGT78	Flavonoid 3-O-glucosyltransferase	Almeida et al., 2007
ant3OGT	ADK75021	*Freesia hybrid cultivar*	F	UGT78	Anthocyanidin-3-O-glucosyltransferase	Sui et al., 2011

Fla3OGT	Q96493	*Gentiana triflora*	F	UGT78	Flavonoid 3-O-glucosyltransferase	Tanaka et al., 1996
Fla3GT	ACH56523	*Gossypium hirsutum*	F	UGT78	Flavonoid 3-O-glucosyltransferase	Yang et al., 2010
BRONZE-1	CAA33729	*Hordeum vulgare*	F	UGT78	Flavonol 3-O-glucosyltransferase	Wise et al., 1990
ant3GT	BAD83701	*Iris hollandica*	F	UGT78	Anthocyanidin-3-O-glucosyltransferase	Yoshihara et al., 2005
Le3GT-A	BAF49310	*Lobelia erinus*	F	UGT78	Anthocyanidin 3-O-glucosyltransferase	Noda., 2005
Fla3OGT	BAB93000	*Malus domestica*	F	UGT78	Flavonoid 3-O-glucosyltransferase	Honda et al., 2002
UGT78G1	ABI94025	*Medicago truncatula*	F	UGT78	(Iso) Flavonoid glycosyltransferase	Modolo et al., 2007
F3GalTase	Q9SBQ8	*Petunia hybrida*	F	UGT78	Kaempferol-3-O-galactosyltransferase	Miller et al., 1999
SaGT4A	BAD89042	*Solanum culeatissimum*	F	UGT73	Saponin glucosyltransferase	Kohara et al., 2005
UF3GaT	BAA36972	*Vigna mungo*	F	UGT78	Flavonoid 3-O-galactosyltransferase	Mato et al., 1998
F3OGT	ABR24135	*Vitis labrusca*	F	UGT78	Flavonoid 3-O-glucosyltransferase	Hall et al., 2007
VvGT1	AAB81683	*Vitis vinifera*	F	UGT78	Flavonoid 3-O-glucosyltransferase	Ford et al., 1998, Offen et al, 2006
UFGT	P51094	*Vitis vinifera*	F	UGT78	Flavonoid 3-O-glucosyltransferase	Kobayashi et al., 2001
BRONZE-1	CAA31856	*Zea mays*	F	UGT78	Flavonoid 3-O-glucosyltransferase	Furtek et al., 1988
UGT85H2	A6XNC5	*Medicago truncatula*	G	UGT85	(Iso) flavonoid glycosyltransferase	Modolo et al., 2007
UGT85A1	AAF18537	*Arabidopsis thaliana*	G	UGT85	Cytokinin glucosyltransferase	Hou et al, 2004
UGT85A24	AB555732	*Gardenia jasminoides*	G	UGT85	Iridoid-O-glucosyltransferase	Nagatoshi., 2011
UGT85K4	AEO45781	*Manihot esculenta*	G	UGT85	Acetone cyanohydrin-O-glucosyltransferase	Kannangara et al., 2011
UGT85K5	AEO45782	*Manihot esculenta*	G	UGT85	2-hydroxy-2-methylbutyronitrile-O-glucosyltransferase	Kannangara et al., 2011
UGT85A19	ABV68925	*Prunus dulcis*	G	UGT85	Mandelonitrile glucosyltransferase	Franks et al., 2008
UGT85B1	AAF17077	*Sorghum bicolor*	G	UGT85	p-Hydroxymandelonitrile-O-glucosyltransferase	Hansen et al., 2003
UGT85C2	AAR06916	*Stevia rebaudiana*	G	UGT85	Steviol glucosyltransferase	Richman et al., 2005
UGT76C1	BAB10792	*Arabidopsis thaliana*	H	UGT76	Cytokinin glucosyltransferase	Hou et al., 2004
UGT76B1	BAC43564	*Arabidopsis thaliana*	H	UGT76	Isoleucic acid-O-glucosyltransferase	von Saint Paul et al., 2011
UGT76C2	AED90933	*Arabidopsis thaliana*	H	UGT76	Cytokinin-N-glucosyltransferase	Hou et al., 2004
UGT76D1	AEC07843	*Arabidopsis thaliana*	H	UGT76	n.d	Ross et al., 2001
UGT76E1	AED97208	*Arabidopsis thaliana*	H	UGT76	n.d	Ross et al., 2001
DicGT5	BAD52007	*Dianthus caryophyllus*	H	UGT76	Chalcononaringenin 2'-O-glucosyltransferase	Ogata et al., 2004
UGT709A4	BAC80066	*Oryza sativa*	H	UGT76	Isoflavonoid glucosyltransferase	Ko et al., 2007b
UGT76G1	AAR06912	*Stevia rebaudiana*	H	UGT76	Steviol glucosyltransferase	Richman et al., 2005

Bx8	AAL57037	*Zea mays*	H	UGT76	Benzoxazinoid glucosyltransferase	von Rad et al., 2001
Bx9	AAL57038	*Zea mays*	H	UGT76	Benzoxazinoid glucosyltransferase	von Rad et al., 2001
UGT83A1	Q9SGA8	*Arabidopsis thaliana*	I	UGT83	n.d	Ross et al., 2001
UGT87A1	O64732	*Arabidopsis thaliana*	J	UGT87	n.d	Ross et al., 2001
UGT86A1	Q9SJL0	*Arabidopsis thaliana*	K	UGT86	n.d	Ross et al., 2001
UGT74E2	AEE27876	*Arabidopsis thaliana*	L	UGT74	Indole-3-butyric acid glucosyltransferase	Tognetti et al., 2010
UGT74B1	AAC00570	*Arabidopsis thaliana*	L	UGT74	thiohydroximate glucosyltransferase	Grubb et al, 2004
UGT74D1	AEC08580	*Arabidopsis thaliana*	L	UGT74	Jasmonate glucosyltransferase	Jin et al, 2013
UGT74F2	AAB64024	*Arabidopsis thaliana*	L	UGT74	Salicylic acid and anthranilic acid glucosyltransferase	Quiel et al, 2003, Dean et al, 2008
UGT75B1	AEE27854	*Arabidopsis thaliana*	L	UGT75	Indole-3-acetic acid glucosyltransferase	Jackson et al, 2001
UGT75C1	CAB10189	*Arabidopsis thaliana*	L	UGT75	Anthocyanin-5-*O*-glucosyltransferase	Gachon et al, 2005
UGT75D1	CAB10333	*Arabidopsis thaliana*	L	UGT75	Flavonoid glucosyltransferase	Tohge et al, 2005
UGT84A1	AEE83609	*Arabidopsis thaliana*	L	UGT84	Hydroxybenzoate glucosyltransferase	Meissner et al, 2008
UGT84B1	AAB87119	*Arabidopsis thaliana*	L	UGT84	Auxine glucosyltransferase	Jackson et al, 2001
SAD10	ACD03250	*Avena strigosa*	L	UGT74	*N*-methylanthranilate-*O*-glucosyltransferase	Owatworakit et al, 2012
UGT74H6	ACD03261	*Avena strigosa*	L	UGT74	Benzoic acid glucosyltransferase	Owatworakit et al, 2012
UGT84A9a	CAS03354	*Brassica napus*	L	UGT84	Sinapate glucosyltransferase	Mittasch et al, 2010
CsGT45	ACM66950	*Crocus sativus*	L	UGT75	Flavonoid glucosyltransferase	Moraga et al, 2009
CaUGT2	Q6X1C0	*Crocus sativus*	L	UGT74	Crocetin-O-glucosyltransferase	Moraga et al., 2004
UGTCs2	AAP94878	*Crocus sativus*	L	UGT74	Crocetin glucosyltransferase	Moraga et al., 2004
EpGT-1	BAD90934	*Eucalyptus perriniana*	L	UGT75	Monoterpene glucosyltransferase	Nagashima et al., 2005
EpGT-2	BAD90935	*Eucalyptus perriniana*	L	UGT75	Monoterpene glucosyltransferase	Nagashima et al., 2005
5GT-A	BAF49285	*Eustoma grandiflorum*	L	UGT75	Anthocyanin 5-*O*-glucosyltransferase	Noda et al., 2002
FaGT2	Q66PF4	*Fragaria ananassa*	L	UGT84	Cinnamate glucosyltransferase	Landmann et al., 2007
UGT75L6	BAM28984	*Gardenia jasminoides*	L	UGT75	Crocetin glucosyltransferase	Nagatoshi et al., 2012
5GT7	BAG32255	*Gentiana triflora*	L	UGT75	Anthocyanin 5-*O*-glucosyltransferase	Nakatsuka et al., 2008
HGT8	BAA36423	*Glandularia x hybrida*	L	UGT75	Anthocyanin 5-*O*-glucosyltransferase	Yamazaki et al., 1999
GgSGT	BAG14302	*Gomphrena globosa*	L	UGT84	Sinapate glucosyltransferase	Matsuba et al., 2008
UGT13248	ADC92550	*Hordeum vulgare*	L	UGT84	Deoxynivalenol-3-O-glucosyltransferase	Schweiger et al., 2010
ant5GT	BAD06874	*Iris hollandica*	L	UGT75	Anthocyanin 5-*O*-glucosyltransferase	Imayama et al., 2004

UGT75L4	ABL85474	*Maclura pomifera*	L	UGT75	Dihydrokaempferol glucosyltransferase	Tian et al., 2006
UGT74M1	ABK76266	*Medicago truncatula*	L	UGT74	Triterpenoid-28-*O*-glucosyltransferase	Meesapyodsuk et al, 2007
ToGT2	AAB36652	*Nicotiana tabacum*	L	UGT75	Coumarins glucosyltransferase	Fraissinet-Tachet, et al., 1998
SAGTase	AAF61647	*Nicotiana tabacum*	L	UGT74	Salicylic acid glucosyltransferase	Lee & Raskin., 1999
PF3R4	BAA36421	*Perilla frutescens*	L	UGT75	Anthocyanin 5-O-glucosyltransferase	Yamazaki et al., 1999
F7GT	AAY27090	*Pyrus communis*	L	UGT75	Flavonoid 7-O-glucosyltransferase	Fischer et al., 2005
Gtsatom	CAI62049	*Solanum lycopersicum*	L	UGT74	Salicylic acid xylosyltransferase	Tarraga et al., 2005
UGT74G1	AAR6920	*Stevia rebaudiana*	L	UGT74	Steviol glucosyltransferase	Richman et al., 2005
ResOGT	ABH03018	*Vitis labrusca*	L	UGT84	Resveratrol-O-glucosyltransferase	Hall & De Luca., 2007
ZmIAGT	AAA59054	*Zea mays*	L	UGT74	Indole-3-acetic acid glucosyltransferase	Szerzen et al., 1994
UGT92A1	Q9LXV0	*Arabidopsis thaliana*	M	UGT92	n.d	Ross et al., 2001
DOPA5GT	BAD91804	*Celosia cristata*	M	UGT92	Cyclo-DOPA-5-O-glucronyltransferase	Sasaki et al., 2005
cDOPA5GT	BAD91803	*Mirabilis jalapa*	M	UGT92	Cyclo-DOPA-5-O-glucosyltransferase	Sasaki et al., 2004
UGT82A1	Q9LHJ2	*Arabidopsis thaliana*	N	UGT82	n.d	Ross et al., 2001
ZOG1	AAD04166	*Phaseolus lunatus*	O	UGT93	Zeatin-O-glucosyltransferase	Martin et al, 1999
ZOX1	AAD51778	*Phaseolus vulgaris*	O	UGT93	Zeatin-O-xylosyltransferase	Martin et al, 1999
cisZog1	AAK53551	*Zea mays*	O	UGT93	cis-Zeatin O-glucosyltransferase	Martin et al., 2001a
cisZog2	AAL92460	*Zea mays*	O	UGT93	cis-Zeatin O-glucosyltransferase	Veach et al., 2003

Supplementary data S.3: Table of UGTs used in Fig. 3.1. Accession numbers refers to GenBank or UniProt (in case of sequences not referenced in GenBank). Literature is referring to publication of functional analysis of UGTs.

Supplementary data S.4: Protein sequences of selected UGTs. PSPG motifs are highlighted in yellow and catalytic residues in green.

SAD10

MGAEWEHVSDIHVLLLPYPVQG<mark>H</mark>INPMLQFGKRLAHIGGVGVRCTLAITP
YLLRQCQDPCPGAVHLVEISDGFDSAGFEEVGDVAAYLAGMESAGSRTLD
ELLRSEAEKGRPIHAVVY<mark>D</mark>AFLQPWVPRVARLHGAACVSFFTQAAAVNVA
YSRRVGKIEEGLPAGFEAEDLPTFLTLPLPYQDMLLSQFVGLDAVDHVLV
NSFHELQPQESAYMESTWGAKTVGPTVPSAYLDKRITDDVSYGFHLYTPM
TATTKAWLDAQPPRSVTYVSFGSMATPGPTEMAEMAEGLHSSGKAFLWVV
RASEASKIPDGFQERVGGRGLVVT<mark>WVAQLEVLAHGAIGCFVTHCGWNSTM
EALGAGVPMVAVPQWSDQ</mark>PTNAKFVEDVWCVGVRARRDPEGVVRREELER
CIREVTGDDKYACNALDWKEKSKRAMSQGGSSDMNITEFLQALRRSRKSY
EAKPIEPLLVGLDA

AsGT3i21

MASTTTATRSSSSSSSRSKKLRVLLIPFFATS<mark>H</mark>IEPFTDLAIRLAAAGSPSVAVEA
TVAVTPANVSIVQSLLERHYGRQHDAAAESTIPVKIATYPFPAVDGLPRGVEN
LGRAAAADSWRIDVAAFSDTLMRPVQEALVREQAPDALVT<mark>D</mark>VHFVWNVRV
AAELGVPCVTFKVVGAFSSLAMRHLALVADVASSDPDVAVVPRFPGLPVRIP
RTELPEFLRKKQEVDYSTTNTFYAAQAACFGLAVNTSSDLEQEYCELHMREG
YVKRAYFIGPVSLRPSPSLDAVGDSQHCVDWLDSKPARSVVYLCFGSFAPVS
EAQLQELALGLEASGESFLWVVRSQEWTPPEGWEERVGDRGMVVTA<mark>WAPQ
TAILGHPAVGAFVTHCGWNSVLETVAAGVPVLTWPMVFEQ</mark>FITERLVTDVLG
IGQRLWPHGAGIRSTRHIENEIVPAEAVARALMAFMCPGGPGDSARNRVMRL
AAKAHAAMAEGGSSHRDLRRLVDDLLEARGAAVAGGPKSVQG

AsGT11i11

MASNDNVPTAVTSSINKKLRVLLIPILATS<mark>H</mark>IGPFTELAISLAATNDAVE
ATVAVTPANVSIVQSMLEHRGGHSVKVATYPFPAVDGLPEGVENFGSAAT
PEQSMCIMVATKSEALTRPVHETLIRSQSPDAVVT<mark>D</mark>MTFLWNSGIAAELG
VPCVVFSVMGAFSMLAMHHLEDAGVDRDDQDDDDDDAVVEVPGFPGPPI
RIPRTELPGFLRRPDYSITNLFISLKAANCFGLAMNTSSELEKQYCELYT
TPPEEGGGGLRRAYFLGPLALALPPPISSSSSSSSDCCSIMAWLDSKPSR
SVVYVSFGSMAHVKDVQLDELALGLETSGISFLWVVRGREEWSPPKGWEA
RVQDRGFIIRA<mark>WAPQISILGHHAAGAFVTQCGWNSVLETVAAAVPMLTWP
LAFEQ</mark>FITERLVTDVLGIGVRLWPDGAGLRSESYQEHEVIPRQDVARALV
EFMRPAAGGPSSIRDMARTKLMDLSAKLHAAVAQGGSSHRDLHRLVDDLL
MEAAAKRPRT

AsGT14h20

MAHTETTEKSMILMSSAAPPPPPHFVLIPLIGQGHTIPMVDLAYLLVERG
ARVSLVTTPVNAARLQGVADRARRAMQPLEIVELPFPPADDGLSPGSANV
DNFLRLFLDLYRLAGPLEAYLRALPRRPSCIISDSCNPWTAGVARSVGVP
RLFFHVASCFYSLCKLKAATHGLLLHDGNKDDAYVEVPGMPVRVEVTKDS
WSSSYTTPEWEAFVEDARDGMRTADGAVLNTFLGLEGQFVKCFEAALGKP
VWALGPFFLNNRDEDAVATRGDKDKPSAVDQDAVTAWLDAMDESAVTYVS
FGSLVRMPPEQLYEVGHGLVDSGKPFVWVVKESETASPEAREWLQALEAR
TAGRGLVVRGWVSQLAILSHRAIGGFVTHCGWNSLLESVAHGVPVVTWPH
FGDQFLNEQLVVEVLGVGVPVRGAAGPVVPVVREHIERAVSELMGGGAVA
QERRRKCKEFGERAHTAVAKGGSSHENLTQLVHSFVRSGSTEQQTTQDRN
C

AsGT14h21

MAPTETAAPPPPPPHFVLVPLIGQGHTIPMGDLACLLAERGARVSLVTTP
VNAARLQGVADRARRARLPLEIVELPLPPADDGLPPGGENSDSIIRLLLA
LYRLAGPLEAYVRALPWRPSCIISDSCNPWMAGVARSVGVPRLFFNGPSC
FYSLCSHNVARHGLLHDGEGEGERDAYVVTGVPVRVEMTKDTWSAALLTC
MPKWEAFLQDVREGMRTADGAVVNTFLDLEEQFVACYRTALGKPVWALGP
FFLGNRDEEAVAARGGKDKPSAVAQSAVTAWLETMDQSTVTYVCFGSFAR
MLPKQLYEVGHGLEDSGKPFLLALKESETALPEAQEWLQALEARTAGKGL
VVRGWAPQLAILSHRAVGGFVTHCGWNSLLESVAHGVPVVTWPHSGDQFL
NERLAIEVLGVGAPVRGAVPVTPFDESKAVAPVLRGHIAEAVSELMGGG
AVARERRRKCKEYGERAHAAIAKGGSSHENLTQLLQSFMRSGSKEQ

AsGT16f23

MTFARGNVHSGSASAHFVLVPMMAQGHTIPMTDMARLLAEHGAQVSFVTT
PVNASRLAGFIADVEAAGLAIRFVELHFPTTEFGLPDGCENLDLIQAKGL
FLNFMEACAALREPLMAHLREQHQLSPPSCIISDMMHWWTGDIARELGIP
RLTFIGFCGFSSLVRYIISQNNLLENMTDENELITIPGFPTHLELTKAKC
PGSLCVPGMEKIREKMIEEELRSDGEVINSFQELETVYIESFEQVAKKKA
WTVGPMCLCHRDSNTMAARGSKASMDEAQCLQWLDSMKPGSVIFVSFGSL
AATTPQQLVELGLGLEASKKPFIWVIKAGPKFPEVEEWLADGFEERVKDR
GMIIRGWAPQMMILWHQAIGGFMTHCGWNSTVEGICAGVPMITWPHFAEQ
FLNEKLVVDVLKTGLEVGVKGVTQWGNTEQEVMVTRDAVETAVYTLMGEG
KAAEELRMRAKHYAIKARRAFDEEGSSYNNVRLLIQEMGNNTNACG

AsGT16h6

MASRQYHVVMVPYPAQSHVAPLMQLARLLHARGAHVTFVHTQFNYRRLVD
AKGEAAVRPSSSTGFCVEVIDDGLSLSVQQHDVAAVVDALRRNCQGPFRA
LLRKLSSAMPPVTTVVADTVMTFAATEAREAGIPDVGFFTASACGLMGYF
QFGELIKRGLVPLQDASCLATPLHWVPGMNHMRLKDMPSFCHTTDPDDTM
VAATLEQMNTALGAKAIVLNTFYELEKDVVDGLAAFFPPLYTVGPLAEVD
SGGSDSLLGAIDISIWQEDAQCLAWLDDKKASSVVYVNFGSIHVMTAAQL
REFALGLASCGFPFLWIKRPDVVVDGEEDAVLPEEFLAAVARGAGLVVPW
CAQPAVLKHPAVGLFVTHCGWNSLLEAAAAGMPLLCWPLFAEQTTNCRQV
CECWGNGAEIPKEVEHGAVSALVREMMEGELGREKRAKAAEWKAAAQTAI
VEGGSSCRSVDRLVEDILLIPSQRK

AsGT21p16

MGSMEQKPHAVCVPFPAQG==H==ITPMLKVAKLLHARGFHVTFVLTDYNYSRL
LRSRGAAAFDGCPGFDFTSIPDGLPPSDAEATQDIPALCRSTMTSCLPHV
RALLARLNGPASAVPPVTCLLC==D==ACMSFAYDAAKEIGLPCAGLWTASGCG
FMAYNYYKNLVEQGIVPLKDQAQLTDGYLDTVVHGVPGVCDGFQLRDFPD
FIRTTDPDDIMLNFLIRETARAASLPDAVIINSFDDLEQRELHAMRAILP
PVCALGPLLLHVRRLVHKGSPLDVAVQSNLWKEQDGLLDWLDGRPPRSVV
YVNYGSITVMTNEQMLEFAWGLANSGYPFLWNVRPDLVKGDAAVLPPEFS
AAIEGRGLLTT==WCPQEKVIVHEAVGVFLTHSGWNSTLESLCAGVPMLSWP==
==FFAEQ==QTNCRYKRTEWGVGMEIGGEVRRAEVAAKIQEAMEGEKGKEMRRR
AAEWKEKAARATLPGGAAEANLDKLIHVLHGKTGQALKRV

AsGT24i2

MAVKDEQQSPLHILLFPFLAPG==H==LIPIADMAALFASRGVRCTILTTPVNA
AIIRSAVDRANDAFRGSDCPAIDISVVPFPDVGLPPGVENGNALTSPADR
LKFFQAVAELREPFDRFLADNHPDAVVS==D==SFFHWSTDAAAEHGVPRLGFLGS
SMFAGSCNESTLHNNPLETAADDPDALVSLPGLPHRVELRRSQTMDPKKRPD
HWALLESVNAADQKSFGEVFNSFHELEPDYVEHYQTTLGRRTWLVGPVALA
SKDMAGRGSTSARSPDADSCLRWLDTKQPGSVVYVSFGTLIRFSPAELHELA
RGLDLSGKNFVWVLGRAGPDSSEWMPQGFADLITPRGDRGFIIRG==WAPQMLI==
==LNHRALGGFVTHCGWNSTLESVSAGVPMVTWPRFADQ==FQNEKLIVEVLKVG
VSIGAKDYGSGIENHDVIRGEVIAESIGKLMGSSEESDAIQRKAKDLGAEARS
AVENGGSSYNDVGRLMDELMARRSSVKVGEDIIPTNDGL

AsGT25n16

MAGMAPLAKTFVLYPSLGVG==H==LNPMVELAKFLVRQGHNVIVAVADPPDSD
AVSADAVARLSAANPCIDFRRLPAPPNPDPAAHPLQRILDTLRLCNPVLR
DFLRSLPGPGAHALLL==D==MFCVHALDVAAELALPAYFFFVSPAGALAVLLN
LPHSYPEMPSFKDMGHQALVRFPGMPPFRAVDMPQGMHDKDSDLTKGLLY
QFSRIPEGRGVLVNTLDWLEPTALRALGDGVCVPGRPTPPVFCIGLLVDG
GYGEKSRPDGGANKCLAWLDKQPHRSVVFLCFGSQGAFSAAQLKEIALGL
ESSGHRFLWAVRSPPEQQGEPDLEGLLPAGFLERTRDRGMVLAD==WVPQAQ==
==VLRHEAVGAFVTHGGWNSAMEAIMSGLPMICWPLYAEQ==ALNKVFMVDEMK
IAVEVAGYEEGMVKAEEVEAKVRLLMETEEGRKLREMLVVARKMALDANA
KGGSSQVAFAKFLCDLENSTST

AsGT27f7

MGTLSELHFVLVPLAAQG==H==IIPMVDVARLIAARGPRVTVVTTPVNAARNR
ATVDGARRAGLPLELVELPFSGPEHGLPEGMEAVDQLTAGAHELGMFLKF
FQAIWNLAGPLEEYIRALPRPPICLVA==D==SCNPWVAAVCERLGIPRLVMHC
PSAYFQLTVNCLITHGVYDRVEEMEPFEVPGFPVRAVGMKATMRGFFQYP
GAEKEYRDTLDAEATADGLLFNTFRGIEGTFLDAYAAALGKPTWAVGPTC
ASSTMADDADSKAGRGNTADVDAGHIVSWLDARPPASVLYVSFGSIAQLT
AKQLANLARGLEASGRPFVWAIKEAKGDAAVRALLDEEGFEARVKDRALL
VRG==WAPQVTILSHPAVGGFLTHCGWNGTLETLSLGVPTLTWPTIADQ==FCS
EQLLVDVLGVGVRSGAKLPAWYLPTEAEGVQVESGDVEKAVAELMGDTPE
AAARRSRAKELAAKARTAMEEGGSSYSDLTDMIRYVSELSRKRSLEIDAM
PFAAAELGSNKGEKIEADAALSVQS

AsGT16525

MAPRPTVVLIPSWGSG**H**FMSALEAGKRLLAASGGAFSLTVLVMHAPTQAKA
SEVEGHVRREAASGLDIRFLQLPAVEHPIDCVDPVEFASRYTQLHGPHVKAAI
ASLAPSFRAAAVVV**D**LFLTALFDVAHELAVPAYVYFASPAAFLALMLRLPAL
REDLTGPGFEEMEGTVDVPGLPPVPPSYMPDCLVRRKIQSYDWFEYHGRRFM
EARGLIVNTTIELEASVLAAIADGRCVPGCPVPALHAIGPVIWFDPKDDDQRR
HECMRWLDDQPPASVVFLCFGSMGSLDAAQVREVAAGLERSGHRFLWVLR
GPPVAGTRFPTDANLDELLPEGFLDATAGRGLVWPA<mark>WAPQREILSHAAVGGF</mark>
<mark>VTHCGWNSVLESLWFGVPLVPWPLYGEQ</mark>HLNAFELVAGVGVAVALEMDR
KKGFFVEAAELERVVRSLMNGGSEEGRKARTKASETSAEFRRAIGEGGSSCA
ALQRLVGEILDLPVGR

AsGT02436

METSAKPHFVLVPWIGSIS**H**IVPMTDIGCLLASHGASVAIITTPANAPLVQSRV
DRVTPRGAVIGVTTIPFPTAEAGLPDGCERLDLVRSPAMVPSFFKANKKFGEA
VAQYCRREDAPRRPSCIIA**G**MCNTWTLPVARDLGVPCYIFHGFGAFALLCIDH
LYRQGRHEAIASAEELVDISVLQQFECKILGRQLPPHFLPSTSMGGGLMQEVR
DFDVAVDGVVVNSFDELEHGSAALLAAASGRKVLAVGPVSLCCAPSLDPQG
DDARRCMAWLDGKEANSVVYVSFGSAGSLPPAQLIQLGMALVSCRWPVMW
VIRGADSLPDDVNAWLGENTPDGVADGKCLVVRG<mark>WAPQVAILAHPAVGG</mark>
<mark>FMTHCGWGSTLESVAAGVPMLTWPLFAEQ</mark>FVNEKLIVDVLAIGVSVGVTKPT
ENVLTASKLGSGEAKAEVGAERVEKALERLMDGAGEGEDIRRRAAELKRKA
NAALEKGGSSYNNLENLIDSCA

AsGT05827

MGIESMDSSVALVAVPFPAQG**H**LNQLMHLSLLVASRGLSVHYAAPAAHVRQ
AKSRVHGWDAKALASIHFHDLDVPTFKSPDPDPAAASPFPSHLLPMWETYSA
AARVPLASLLERLSATHRRVVVVY**D**HMNSFAAAEAARVDGEAYGLVCVAIS
NHLAWMPDGHQLLRDRGLRSVPMDACMSKEFVEYMARVTTEAEGAGFLM
NTCRALEGEFLDAVAEIPDIKRQKRFAVGPLNPLLPLATEPDVVTTATARHDC
MRWLDAQPPASVLYVSFGTTSSFLPEQIAELAAAIKGSRQRFIWVLRDADRA
DIFAGNSGGDSRPRYEKLLSEEQAQGTGLVITG<mark>WAPQLEILAHGATAAFMSH</mark>
<mark>CGWNSTMESLSHGKPILAWPMHSDQ</mark>PWDAELLCNYLRVAMLVRPWEKHGE
VVAAGAIQEVIEKAMLSEKGTALRQRAKLLGEAVRAAVADNGSSTKDLHDF
VAHITRI

AsGT23586A

MKQTVVLYPGAGGS**H**VAAMTELANVFLKHGYDVTMVLVEPPFKSSDSGAT
AIERIAASNPSISFHVLPPLPPPDFAAAGNKNPFVLMFQLLLEYNELLEAFLRSI
PRKRLHSVVL**D**MFCIHALDVCVKLGVPVYTFFASGASCLSVLTQFPALIAGRQ
TGLKEIGDTPLDFLGVPPMLASHIIKELLEHPEDEMCKILTNMWKRNTETMGV
LVNTFESLESRAVQSLRDPLCVPGRILPPIYCVGPLVGEGAKDGDGAERNECL
AWLDSQPDRSVVFLCFGSKGTVSAEQLKEIAVGLERSGQRFLWSVRTPAGSQ
DAKKYLEVRAEPDLDELMPEGFLERTKDKGLVIKS<mark>WAPQVDVLRHRATGAF</mark>
<mark>VTHCGWNSVLEAVSAGVPMLCWPLEAEQ</mark>KMNKVCMTEDMGVAVELDGY
MAGFVKADEVEAKVRLVIEGEDGRQLRARVAARKEEAEAALEEGGASRAAF
VQFLLDVENIGEQVRE

AsGT23586B

MKQTVVLYPGAGVS<mark>H</mark>VGPMTELANVFLKHGYDVTMVLVEPPFKSSDSGATA
IERIAASNPSISFHVLPPLPPPDFAAAGNKNPFVLMFQLLLEYNELLEAFLRSIP
RKRLHSVVL<mark>D</mark>MFCIHALDVCVKLGVPVYTFFASGASCLSVLTQFPALIAGRQT
GLKEIGDTPLDFLGVPPMPASHLIKEMLEHPEEELCKEITNIWKRNTETMGVL
VNTFESLESRAVQSLRDPLCVPGRTLPPIYCVGPLVGEGAKDGGGTERNKCLA
WLDSQPDRSVVFLCFGSKGTLSAEQLKEIAVGLERSGQRFLWSVRTPAGSDD
AMKYLEVRPEPDLDALMPEGFLERTKDRGLVVKS<mark>WAPQVDVLRHRATGAF
VTHCGWNSVLEAVSAGVPMLCWPLEAEQ</mark>KMNKVCMTEDMGVAVELDGYR
TGFVKADEVEAKVRLVLESEEGRQLRARVAARKEEAEAALEEGGASRAAFV
QFLLDVDSIGEQVRE

AsGT23781

MEAPHFVFVPLMAQG<mark>H</mark>IIPAVDTALLLATHGALCTVVATPSTAARVRPTIDSA
RRSGLQVSLVDFPLDYAAVGLPDGVPGGADNMDNVPPEYMLAYYSAIALLC
GPIETYLRAHAPRPPTCVVS<mark>D</mark>FCHPWTTKLAASLGVPRLSFFSMCAFCVLCQH
NVERFNAYDGVLDPNQPVVVPGLEKRFEVTMAQAPGFFRGWPGWEKFADD
VERARVDADGVVMNTFEEMEPEYVAGYAAARGMKVWTVGPVSLYHQHAR
TLAARGATASIDTDDCVRWLDGKDPGSVVYVSFGSIVHADPKQIMELGLGLE
ASGYPFVWVVKGAERHDEAALAFLRGLEERVAGRGLLVWG<mark>WAPQALILSH
RAAGAFVTHCGWNSTLEAVAAGLPVVTWPHFTDQ</mark>FLNEKLAVEVLGIGVSV
GVKEPVMYQVDKKEIVVGRSAVEAAVRSAMDGGEDGQERRRRARELAVKA
RAAVSEGGSSHTNIRDLVKCFGVGASTQDAAE

Arabidopsis thaliana

Group D

Name	GenBank
UGT73B1	AL021961
UGT73B2	AY035164
UGT73B2	AL021961
UGT73B3	AL021961
UGT73B4	AC006248
UGT73B5	AC006248
UGT73C1	AC006282
UGT73C2	AC006282
UGT73C3	AC006282
UGT73C4	AC006282
UGT73C5	AC006282
UGT73C6	AC006282
UGT73C7	AL132958
UGT73D1	AL132958

Oryza sativa

Group D

| Name	Reference sequence
RUGT-5	NP_001043481
Os01g0175700	NP_001172201
Os01g0176000	NP_001042177
Os01g0176100	NP_001042178
Os01g0176200	NP_001042179
Os01g0638000	NP_001043671
Os01g0638600	NP_001043672
Os02g0206100	NP_001046251
Os02g0206400	NP_001046252
Os02g0206700	NP_001046253
Os03g0212000	NP_001173311
Os03g0358800	NP_001050148
Os03g0666600	NP_001050847
Os03g0745100	NP_001051242
Os03g0808200	NP_001051649
Os04g0305700	NP_001052417
Os04g0523600	NP_001053346
Os04g0523700	NP_001053347
Os05g0177500	NP_001054797
Os05g0499600	NP_001055958
Os05g0499800	NP_001055959
Os05g0500000	NP_001055960
Os06g0187500	NP_001057016
Os06g0590700	NP_001057969
Os09g0329200	NP_001062897
Os09g0329700	NP_001062898
Os09g0379300	NP_001175809
Os10g0178500	NP_001064247

Supplementary data S.5: Complete collections of group D (UGT73) sequences from *Arabidopsis thaliana* **and** *Oryza sativa*.
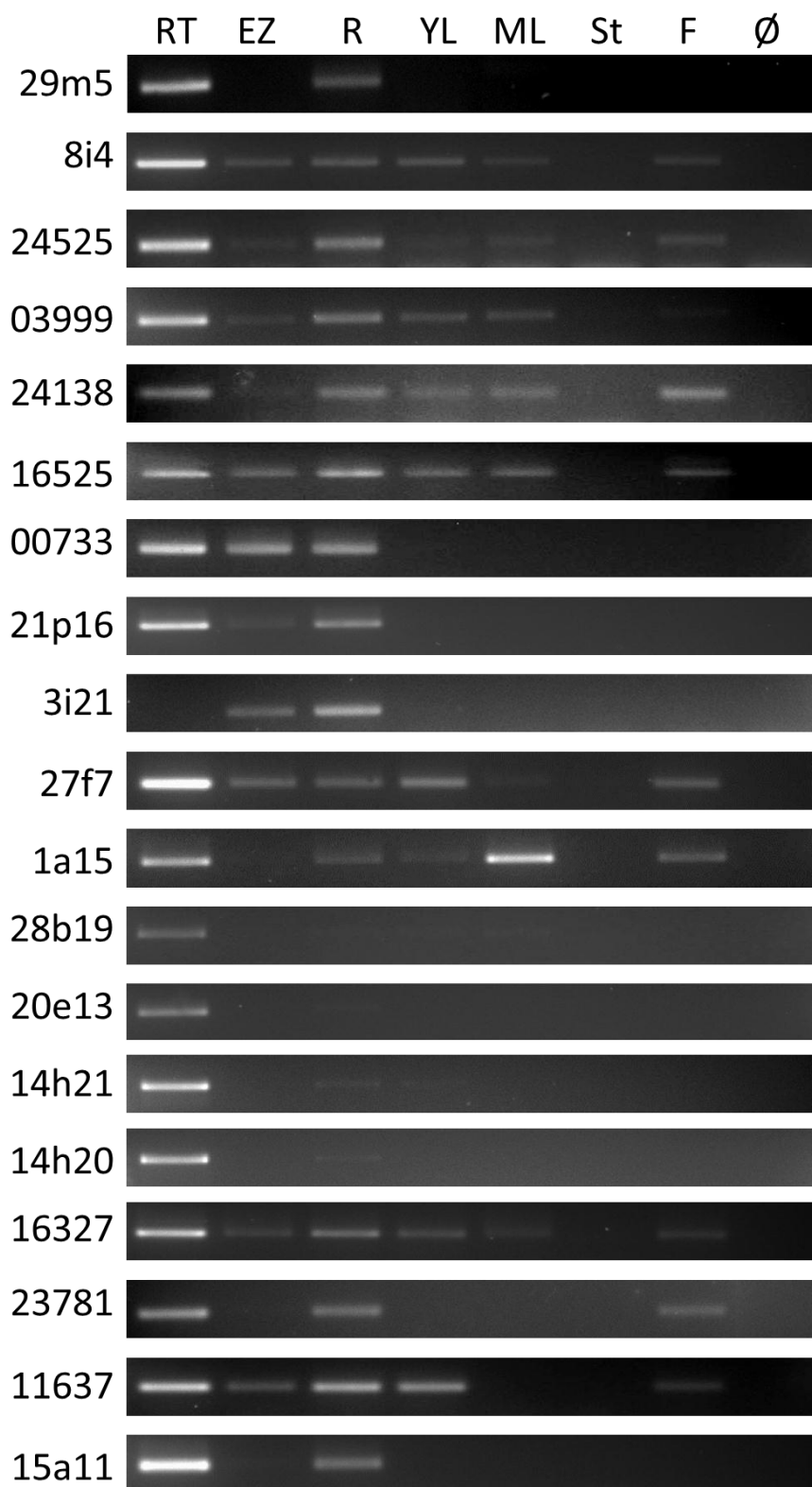
Supplementary data S.6: Expression analysis of UGT genes in *A. strigosa* tissues. Expression analysis was conducted using mRNA-reverse transcription-PCR (RT-PCR) technique. *A. strigosa* tissues used were from 3-days-old seedlings (RT: root tip, RE: root elongation zone; R: entire young root, YL: young leaf) or tissues of flowering plants (ML: mature leaf, St: stalk and F: flower).

Supplementary data S.7: H¹ NMR spectrum of TCP-Glc.
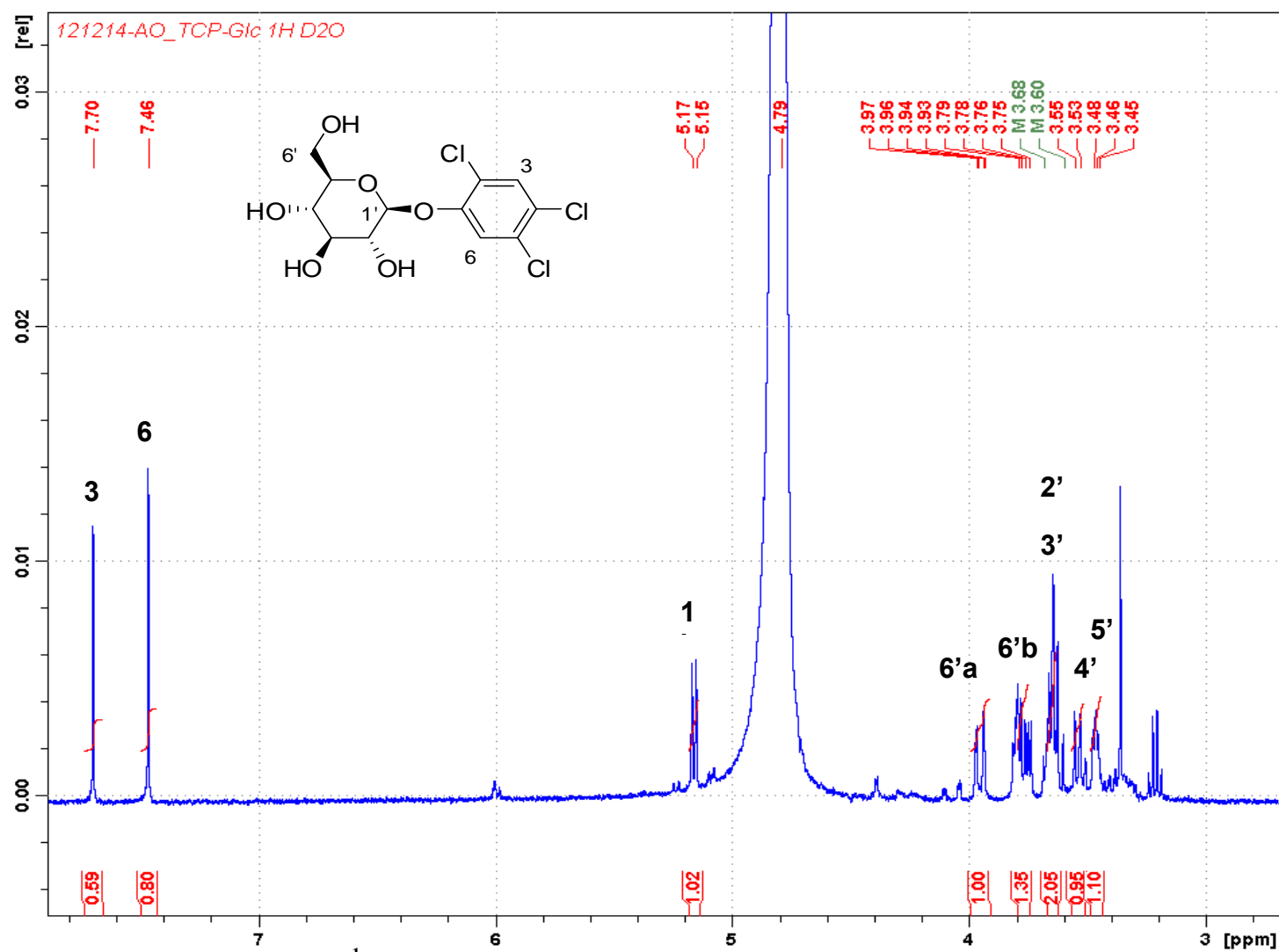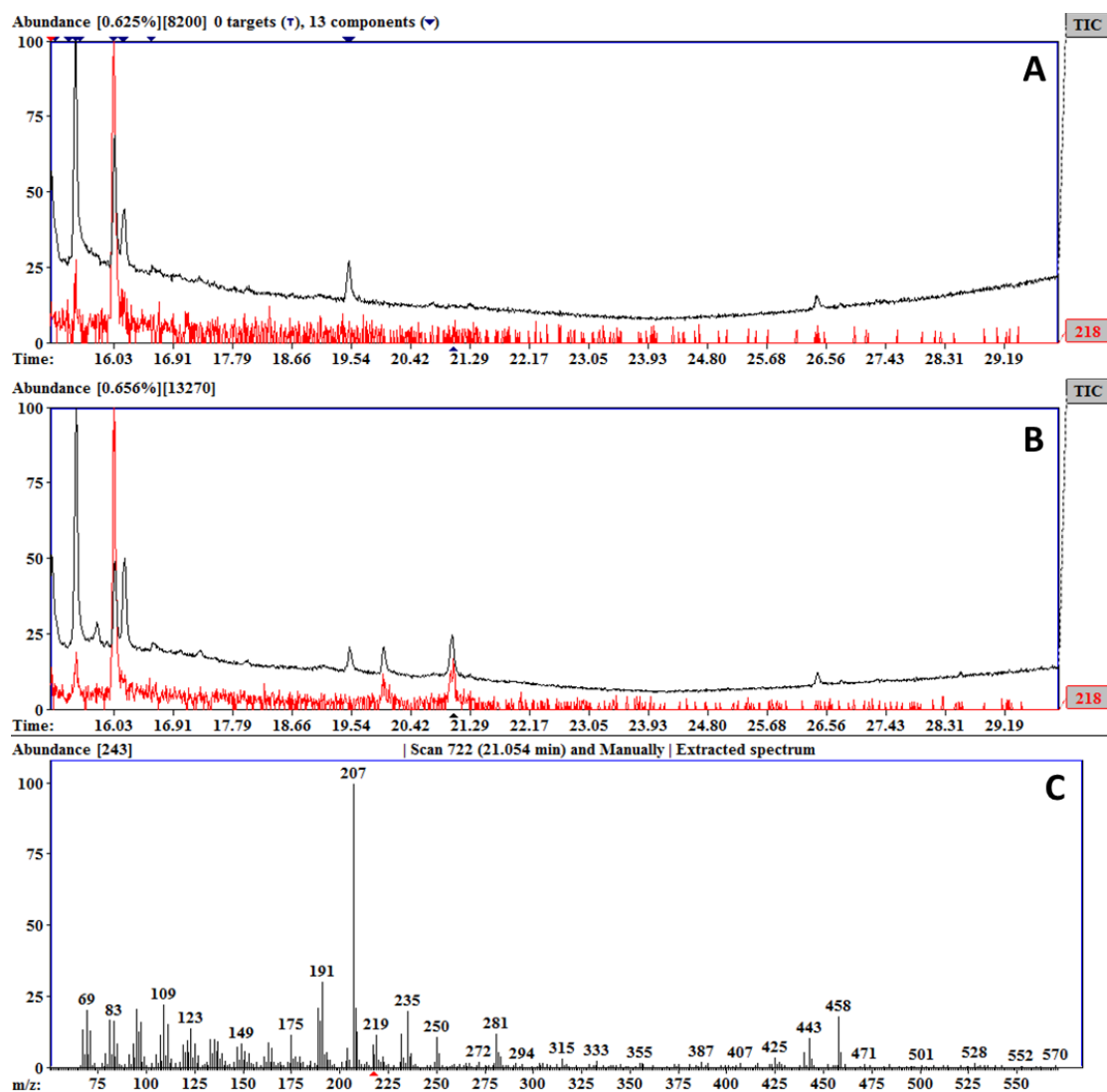
Supplementary data S.8: Consumption of 12,13-epoxy-16-hydroxy-β-amyrin in the presence of UGT73C10. GC-MS analysis of leaf extracts from *N. benthamiana* tissues expressing SAD1 and SAD2 with UGT73C10 (**A**) or with the GFP control (**B**). Total ion chromatogram and selected m/z 218 (base peak at m/z 218 is typical for triterpenes that contain a C-12/C-13 double following a retro-Diels-Alder fragmentation) are displayed in panels **A** and **B**. and. The peak of 12,13-epoxy-16-hydroxy-β-amyrin eluting at Rt: 21.0 min is absent in tissues expressing UGT73C10 (β-amyrin elutes at Rt: 16.0 min) suggesting the glycosyltransferase is able to use the SAD2 product as acceptor and consumes the entire SAD2 product. **C.** The extracted ion chromatogram from 12,13-epoxy-16-hydroxy-β-amyrin peak is consistent with Geisler et al. (2013), notably the fragment m/z 219 is prominent over m/z 218 due to modification of retro-Diels-Alder fragmentation imputable to the epoxide group.
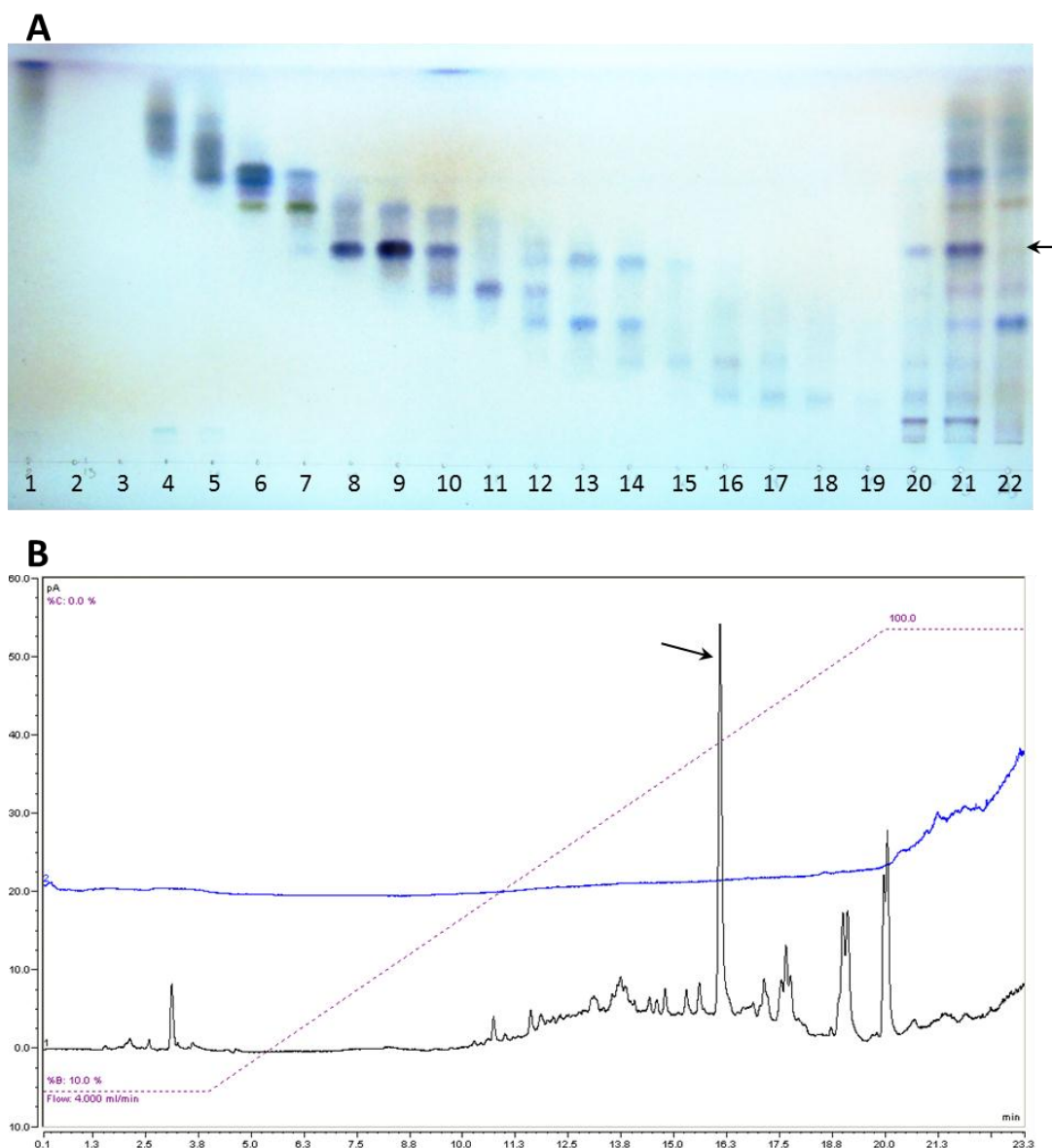
Supplementary data S.9: purification of SAD1-SAD2-UGT73C10 co-expression product from *N. benthamiana* agroinfiltrated tissues. A. Analytical thin layer chromatography was prepared as detailed in section 2.2.23. Lanes 1-19 are loaded with liquid chromatography fractions containing biological material. Lane 20 is loaded with the organic phase from ethyl acetate partitioning of *N. benthamiana* extracts (sample applied to the chromatographic silica column). Lane 21 and 22 are loaded with crude methanolic extract of SAD1-SAD2-UGT73C10 and SAD1-SAD2-GFP expressing tissues respectively. The targeted compound is indicated by an arrow. **B.** Fractions 7, 8 and 9 were pooled and purified using a reverse phase semi-preparative HPLC coupled with a universal charged aerosol detector. The blue spectrum is from blank injection, the black spectrum is from pooled sample of the three fractions containing the targeted compound (see panel A). The targeted peak indicated by an arrow has been collected.

131216-AO_Oxidised b-amyrin-Glc purified HSQCed CD3CN

8 x CH$_3$
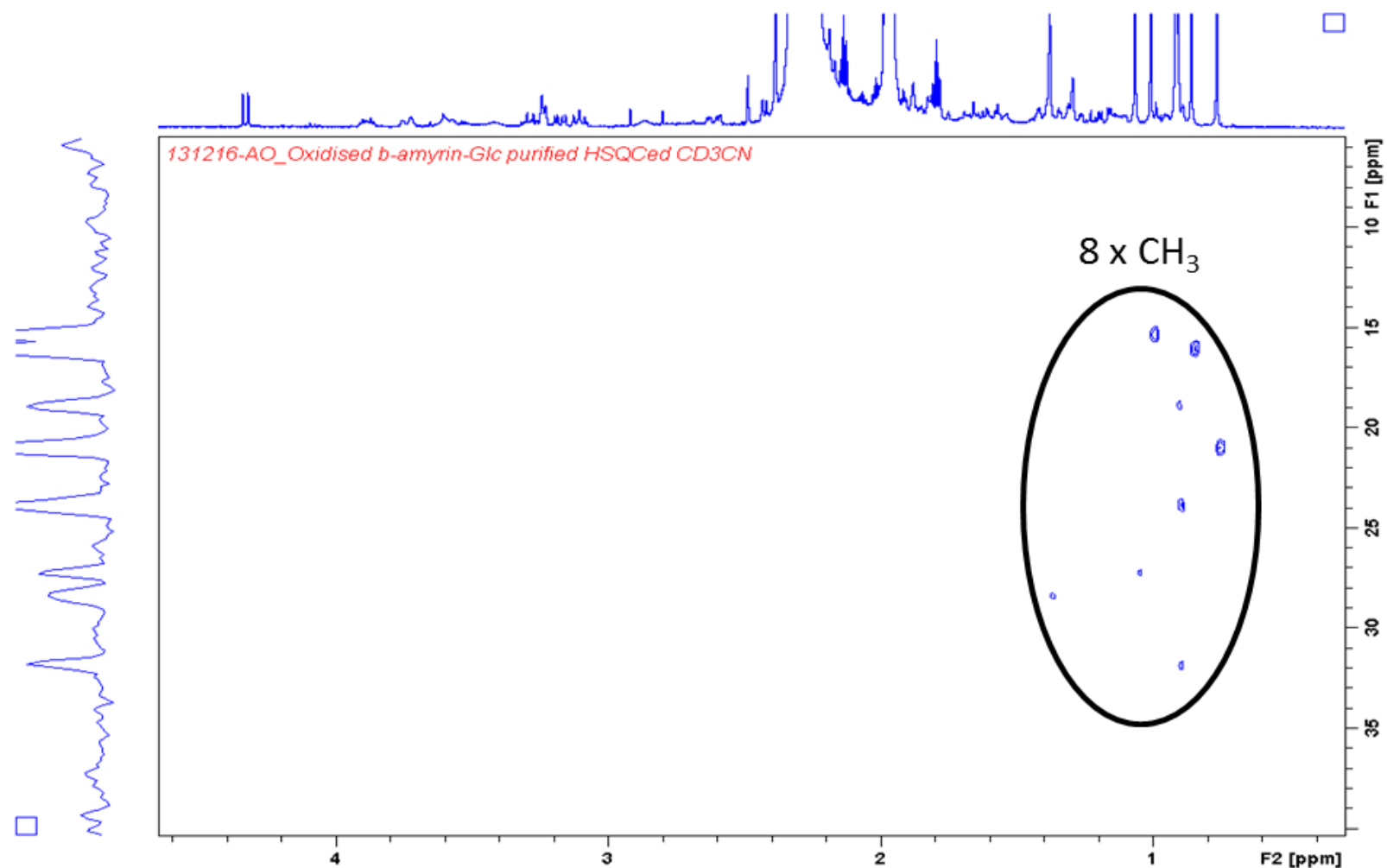
Supplementary data S.10: HSQC analysis of the purified SAD1-SAD2-UGT73C10 co-expression product extracted and purified from agroiniltrated *N. benthamiana*.
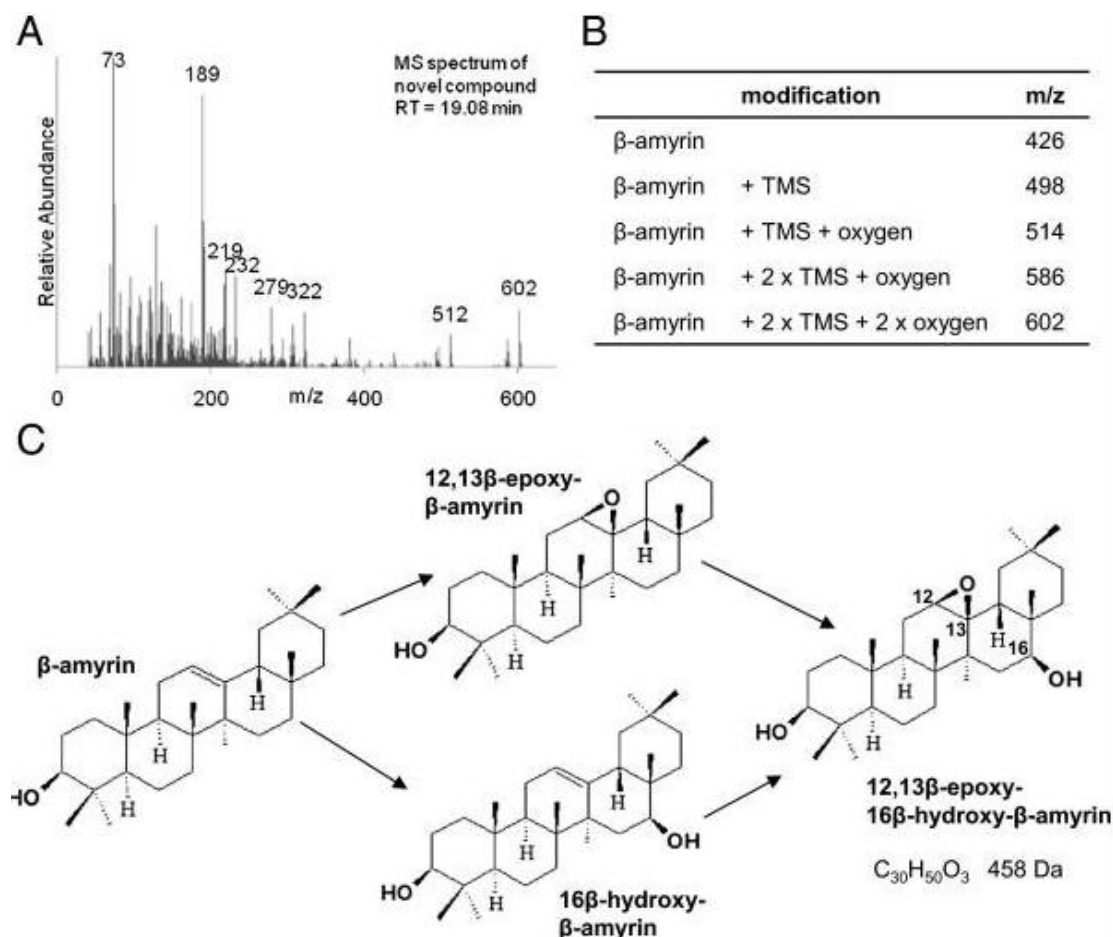
Supplementary data S.11: Identification of the product generated by coexpression of SAD1 and SAD2. (A) Automated mass spectral deconvolution and identification system-extracted ion component spectra of the trimethylsilyated compound at the indicated retention time (RT). The compound has a predicted molecular ion at m/z = 602 that is consistent with the molecular equation C36H66O3Si2. The signal at m/z = 512 is consistent with the loss of C3H9OSi, resulting in the molecular equation C30H56O2Si. (B) Calculated m/z values for β-amyrin derivates, considering trimethylsilyl (TMS) derivatization (+72) and introduction of oxygen atoms (+16). (C) Structure, chemical formula, and molecular mass of 12,13β-epoxy-3β,16β-dihydroxy-oleanane (12,13β-epoxy-16β-hydroxy-β-amyrin). TMS-derivatized 12,13β-epoxy-16β-hydroxy-β-amyrin has a predicted molecular ion at m/z = 602. Potential intermediates in the synthesis of 12,13β-epoxy-16β-hydroxy-β-amyrin from β-amyrin are shown also. This is a copy of the figure 2 from Geisler et al. 2013.

References

Achnine L, Huhman DV, Farag MA, Sumner LW, Blount JW, Dixon RA (2005) Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula. Plant J 41: 875-887

Agati G, Biricolti S, Guidi L, Ferrini F, Fini A, Tattini M (2011) The biosynthesis of flavonoids is enhanced similarly by UV radiation and root zone salinity in L. vulgare leaves. J Plant Physiol 168: 204-212

Armah CN, Mackie AR, Roy C, Price K, Osbourn AE, Bowyer P, Ladha S (1999) The membrane-permeabilizing effect of avenacin A-1 involves the reorganization of bilayer cholesterol. Biophys J 76: 281-290

Augustin JM, Drok S, Shinoda T, Sanmiya K, Nielsen JK, Khakimov B, Olsen CE, Hansen EH, Kuzina V, Ekstrom CT, Hauser T, Bak S (2012) UDP-glycosyltransferases from the UGT73C subfamily in Barbarea vulgaris catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. Plant Physiol 160: 1881-1895

Augustin JM, Kuzina V, Andersen SB, Bak S (2011) Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry 72: 435-457

Baez D, Pino JA, Morales D (2012) Floral scent composition of Plumeria tuberculata analyzed by HS-SPME. Nat Prod Commun 7: 101-102

Bajguz A, Piotrowska A (2009) Conjugates of auxin and cytokinin. Phytochemistry 70: 957-969

Bak S, Beisson F, Bishop G, Hamberger B, Hofer R, Paquette S, Werck-Reichhart D (2011) Cytochromes p450. The Arabidopsis book / American Society of Plant Biologists 9: e0144

Baneyx F (1999) Recombinant protein expression in Escherichia coli. Curr Opin Biotechnol 10: 411-421

Baneyx F, Mujacic M (2004) Recombinant protein folding and misfolding in Escherichia coli. Nature biotechnology 22: 1399-1408

Bedir E, Calis I, Zerbe O, Sticher O (1998) Cyclocephaloside I: A Novel Cycloartane-Type Glycoside from Astragalus microcephalus. Journal of natural products 61: 503–505

Bowles D, Isayenkova J, Lim EK, Poppenberger B (2005) Glycosyltransferases: managers of small molecules. Curr Opin Plant Biol 8: 254-263

Bowles D, Lim EK, Poppenberger B, Vaistij FE (2006) Glycosyltransferases of lipophilic small molecules. Annu Rev Plant Biol 57: 567-597

Bowyer P, Clarke BR, Lunness P, Daniels MJ, Osbourn AE (1995) Host range of a plant pathogenic fungus determined by a saponin detoxifying enzyme. Science 267: 371-374

Brazier-Hicks M, Edwards LA, Edwards R (2007a) Selection of plants for roles in phytoremediation: the importance of glucosylation. Plant Biotechnol J 5: 627-635

Brazier-Hicks M, Evans KM, Gershater MC, Puschmann H, Steel PG, Edwards R (2009) The C-glycosylation of flavonoids in cereals. J Biol Chem 284: 17926-17934

Brazier-Hicks M, Offen WA, Gershater MC, Revett TJ, Lim EK, Bowles DJ, Davies GJ, Edwards R (2007b) Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. Proc Natl Acad Sci U S A 104: 20238-20243

Buer CS, Imin N, Djordjevic MA (2010) Flavonoids: new roles for old molecules. Journal of integrative plant biology 52: 98-111

Buschhaus C, Jetter R (2012) Composition and physiological function of the wax layers coating Arabidopsis leaves: beta-amyrin negatively affects the intracuticular water barrier. Plant Physiol 160: 1120-1129

Cai H, Sale S, Schmid R, Britton RG, Brown K, Steward WP, Gescher AJ (2009) Flavones as colorectal cancer chemopreventive agents--phenol-o-methylation enhances efficacy. Cancer prevention research 2: 743-750

Campbell JA, Davies GJ, Bulone V, Henrissat B (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochemical Journal 326: 929-939

Cane DE, Ikeda H (2012) Exploration and mining of the bacterial terpenome. Acc Chem Res 45: 463-472

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic acids research 37: D233-238

Cao G, Prior RL (1999) Anthocyanins are detected in human plasma after oral administration of an elderberry extract. Clinical chemistry 45: 574-576

Caputi L, Lim EK, Bowles DJ (2008) Discovery of new biocatalysts for the glycosylation of terpenoid scaffolds. Chemistry 14: 6656-6662

Caputi L, Malnoy M, Goremykin V, Nikiforova S, Martens S (2011) A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. Plant J 69: 1030-1042

Caputi L, Rejzek M, Louveau T, O'Neill EC, Hill L, Osbourn A, Field RA (2013) A one-pot enzymatic approach to the O-fluoroglucoside of N-methylanthranilate. Bioorganic & medicinal chemistry 21: 4762-4767

Cartwright AM, Lim EK, Kleanthous C, Bowles DJ (2008) A kinetic analysis of regiospecific glucosylation by two glycosyltransferases of Arabidopsis thaliana: domain swapping to introduce new activities. J Biol Chem 283: 15724-15731

Cavalieri E, Mariotto S, Fabrizi C, de Prati AC, Gottardo R, Leone S, Berra LV, Lauro GM, Ciampa AR, Suzuki H (2004) alpha-Bisabolol, a nontoxic natural compound, strongly induces apoptosis in glioma cells. Biochemical and biophysical research communications 315: 589-594

Chappell J, Wolf F, Proulx J, Cuellar R, Saunders C (1995) Is the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase a Rate-Limiting Step for Isoprenoid Biosynthesis in Plants? Plant Physiol 109: 1337-1343

Chaturvedi P, Misra P, Tuli R (2011) Sterol glycosyltransferases--the enzymes that modify sterols. Applied biochemistry and biotechnology 165: 47-68

Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J 66: 212-229

Chung BKS, Lakshmanan M, Klement M, Mohanty B, Lee DY (2013) Genome-scale in silico modeling and analysis for designing synthetic terpenoid-producing microbial cell factories. Chemical Engineering Science 103: 100-108

Cole D, Edwards R (2000) Secondary metabolism of agrochemicals in plants. Metabolism of agrochemicals in plants: 107-154

Conchie J, Gelman AL, Levvy GA (1968) Inhibition of glycosidases by aldonolactones of corresponding configuration. The specificity of alpha-L-arabinosidase. Biochem J 106: 135-140

Connolly JD, Hill RA (2010) Triterpenoids. Natural product reports 27: 79-132

Cook RJ (2003) Take-all of wheat. Physiological and Molecular Plant Pathology 62: 73–86

Crombie L, Crombie WML, Whiting DA (1984) Structures of the four avenacins, oat root resistance factors to 'take-all' disease. J. Chem. Soc., Chem. Commun: 246-248

Custodio L, Serra H, Nogueira JM, Goncalves S, Romano A (2006) Analysis of the volatiles emitted by whole flowers and isolated flower organs of the carob tree using HS-SPME-GC/MS. Journal of chemical ecology 32: 929-942

D'Auria JC, Gershenzon J (2005) The secondary metabolism of Arabidopsis thaliana: growing like a weed. Curr Opin Plant Biol 8: 308-316

Darra E, Abdel-Azeim S, Manara A, Shoji K, Marechal JD, Mariotto S, Cavalieri E, Perbellini L, Pizza C, Perahia D, Crimi M, Suzuki H (2008) Insight into the apoptosis-inducing action of alpha-bisabolol towards malignant tumor cells: involvement of lipid rafts and Bid. Arch Biochem Biophys 476: 113-123

Daviere JM, Achard P (2013) Gibberellin signaling in plants. Development 140: 1147-1151

Day JA, Saunders EM (2004) Glycosidation of chlorophenols by Lemna minor. Environmental toxicology and chemistry / SETAC 23: 613-620

De Marino S, Borbone N, Gala F, Zollo F, Fico G, Pagiotti R, Iorizzi M (2006) New constituents of sweet Capsicum annuum L. fruits and evaluation of their biological activity. Journal of agricultural and food chemistry 54: 7508-7516

Dinda B, Debnath S, Mohanta BC, Harigaya Y (2010) Naturally occurring triterpenoid saponins. Chemistry & biodiversity 7: 2327-2580

Dixon RA (2001) Natural products and plant disease resistance. Nature 411: 843-847

Enfissi EM, Fraser PD, Lois LM, Boronat A, Schuch W, Bramley PM (2005) Metabolic engineering of the mevalonate and non-mevalonate isopentenyl diphosphate-forming pathways for the production of health-promoting isoprenoids in tomato. Plant Biotechnol J 3: 17-27

Fay P, Duke W (1977) An assessment of allelopathic potential in Avena germ plasm. Weed Science: 224-228

Flores HE, Vivanco JM, Loyola-Vargas VM (1999) 'Radicle' biochemistry: the biology of root-specific metabolism. Trends Plant Sci 4: 220-226

Ford CM, Boss PK, Hoj PB (1998) Cloning and characterization of Vitis vinifera UDP-glucose:flavonoid 3-O-glucosyltransferase, a homologue of the enzyme encoded by the maize Bronze-1 locus that may primarily serve to glucosylate anthocyanidins in vivo. J Biol Chem 273: 9224-9233

Francis G, Kerem Z, Makkar HP, Becker K (2002) The biological action of saponins in animal systems: a review. The British journal of nutrition 88: 587-605

Frydman A, Liberman R, Huhman DV, Carmeli-Weissberg M, Sapir-Mir M, Ophir R, Sumner LW, Eyal Y (2013) The molecular and enzymatic basis of bitter/non-bitter flavor of citrus fruit: evolution of branch-forming rhamnosyl-transferases under domestication. Plant Journal 73: 166-178

Gachon C, Baltz R, Saindrenan P (2004) Over-expression of a scopoletin glucosyltransferase in Nicotiana tabacum leads to precocious lesion formation during the hypersensitive response to tobacco mosaic virus but does not affect virus resistance. Plant Mol Biol 54: 137-146

Gachon CM, Langlois-Meurinne M, Henry Y, Saindrenan P (2005a) Transcriptional co-regulation of secondary metabolism enzymes in Arabidopsis: functional and evolutionary implications. Plant Mol Biol 58: 229-245

Gachon CM, Langlois-Meurinne M, Saindrenan P (2005b) Plant secondary metabolism glycosyltransferases: the emerging functional analysis. Trends Plant Sci 10: 542-549

Gantt RW, Peltier-Pain P, Cournoyer WJ, Thorson JS (2011) Using simple donors to drive the equilibria of glycosyltransferase-catalyzed reactions. Nature chemical biology 7: 685-691

Geisler K, Hughes RK, Sainsbury F, Lomonossoff GP, Rejzek M, Fairhurst S, Olsen CE, Motawia MS, Melton RE, Hemmings AM, Bak S, Osbourn A (2013) Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. Proc Natl Acad Sci U S A 110: 3360-3367

Ghantous A, Gali-Muhtasib H, Vuorela H, Saliba NA, Darwiche N (2010) What made sesquiterpene lactones reach cancer clinical trials? Drug discovery today 15: 668-678

Gonzalez-Lamothe R, Mitchell G, Gattuso M, Diarra MS, Malouin F, Bouarab K (2009) Plant antimicrobial agents and their effects on plant and human pathogens. International journal of molecular sciences 10: 3400-3419

Gopalan A, Reuben SC, Ahmed S, Darvesh AS, Hohmann J, Bishayee A (2012) The health benefits of blackcurrants. Food & function 3: 795-809

Grille S, Zaslawski A, Thiele S, Plat J, Warnecke D (2010) The functions of steryl glycosides come to those who wait: Recent advances in plants, fungi, bacteria and animals. Prog Lipid Res 49: 262-288

Grubb CD, Zipp BJ, Ludwig-Muller J, Masuno MN, Molinski TF, Abel S (2004) Arabidopsis glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. Plant J 40: 893-908

Halliwell CM, Morgan G, Ou CP, Cass AE (2001) Introduction of a (poly)histidine tag in L-lactate dehydrogenase produces a mixture of active and inactive molecules. Analytical biochemistry 295: 257-261

Hansen EH, Osmani SA, Kristensen C, Moller BL, Hansen J (2009) Substrate specificities of family 1 UGTs gained by domain swapping. Phytochemistry 70: 473-482

Hansen KS, Kristensen C, Tattersall DB, Jones PR, Olsen CE, Bak S, Moller BL (2003) The in vitro substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from Sorghum bicolor. Phytochemistry 64: 143-151

Haralampidis K, Bryan G, Qi X, Papadopoulou K, Bakht S, Melton R, Osbourn A (2001) A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. Proc Natl Acad Sci U S A 98: 13431-13436

Harborne J, Baxter H (1999) The Handbook of Natural Flavonoids, vol. 1John Wiley & Sons. Chichester, UK

Hartley JL, Temple GF, Brasch MA (2000) DNA cloning using in vitro site-specific recombination. Genome research 10: 1788-1795

Hartmann T (2007) From waste products to ecochemicals: fifty years research of plant secondary metabolism. Phytochemistry 68: 2831-2846

He XZ, Wang X, Dixon RA (2006) Mutational analysis of the Medicago glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (iso)flavonoid glycosylation. J Biol Chem 281: 34441-34447

Heim KE, Tagliaferro AR, Bobilya DJ (2002) Flavonoid antioxidants: chemistry, metabolism and structure-activity relationships. Journal of Nutritional Biochemistry 13: 572-584

Helmy M, Tomita M, Ishihama Y (2011) OryzaPG-DB: rice proteome database based on shotgun proteogenomics. BMC Plant Biol 11: 63

Hemmerlin A, Harwood JL, Bach TJ (2012) A raison d'etre for two distinct pathways in the early steps of plant isoprenoid biosynthesis? Prog Lipid Res 51: 95-148

Hirotani M, Kuroda R, Suzuki H, Yoshikawa T (2000) Cloning and expression of UDP-glucose: flavonoid 7-O-glucosyltransferase from hairy root cultures of Scutellaria baicalensis. Planta 210: 1006-1013

Hollman PC, Bijsman MN, van Gameren Y, Cnossen EP, de Vries JH, Katan MB (1999) The sugar moiety is a major determinant of the absorption of dietary flavonoid glycosides in man. Free radical research 31: 569-573

Holmberg N, Harker M, Wallace AD, Clayton JC, Gibbard CL, Safford R (2003) Co-expression of N-terminal truncated 3-hydroxy-3-methylglutaryl CoA reductase and C24-sterol methyltransferase type 1 in transgenic tobacco enhances carbon flux towards end-product sterols. Plant J 36: 12-20

Hou B, Lim EK, Higgins GS, Bowles DJ (2004) N-glucosylation of cytokinins by glycosyltransferases of Arabidopsis thaliana. J Biol Chem 279: 47822-47832

Howe GA, Jander G (2008) Plant immunity to insect herbivores. Annu Rev Plant Biol 59: 41-66

Hu Y, Walker S (2002) Remarkable structural similarities between diverse glycosyltransferases. Chemistry & biology 9: 1287-1296

Huffaker A, Kaplan F, Vaughan MM, Dafoe NJ, Ni X, Rocca JR, Alborn HT, Teal PE, Schmelz EA (2011) Novel acidic sesquiterpenoids constitute a dominant class of pathogen-induced phytoalexins in maize. Plant Physiol 156: 2082-2097

Hughes J, Hughes MA (1994) Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons. 5: 41-49

Huhman DV, Sumner LW (2002) Metabolic profiling of saponins in Medicago sativa and Medicago truncatula using HPLC coupled to an electrospray ion-trap mass spectrometer. Phytochemistry 59: 347-360

Husar S, Berthiller F, Fujioka S, Rozhon W, Khan M, Kalaivanan F, Elias L, Higgins GS, Li Y, Schuhmacher R, Krska R, Seto H, Vaistij FE, Bowles D, Poppenberger B (2011) Overexpression of the UGT73C6 alters brassinosteroid glucoside formation in Arabidopsis thaliana. BMC Plant Biol 11: 51

Ikeda T, Udayama M, Okawa M, Arao T, Kinjo J, Nohara T (1998) Partial hydrolysis of soyasaponin I and the hepatoprotective effects of the hydrolytic products. Study of the structure-hepatoprotective relationship of soyasapogenol B analogs. Chemical & pharmaceutical bulletin 46: 359-361

Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R, Tikunov Y, Bovy A, Chikate Y, Singh P, Rogachev I, Beekwilder J, Giri AP, Aharoni A (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science 341: 175-179

Itkin M, Rogachev I, Alkan N, Rosenberg T, Malitsky S, Masini L, Meir S, Iijima Y, Aoki K, de Vos R, Prusky D, Burdman S, Beekwilder J, Aharoni A (2011) GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. Plant Cell 23: 4507-4525

Jin SH, Ma XM, Han P, Wang B, Sun YG, Zhang GZ, Li YJ, Hou BK (2013) UGT74D1 Is a Novel Auxin Glycosyltransferase from Arabidopsis thaliana. PloS one 8: e61705

Jones P, Messner B, Nakajima J, Schaffner AR, Saito K (2003) UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in Arabidopsis thaliana. J Biol Chem 278: 43910-43918

Jones P, Vogt T (2001) Glycosyltransferases in secondary plant metabolism: tranquilizers and stimulant controllers. Planta 213: 164-174

Jorasch P, Warnecke DC, Lindner B, Zahringer U, Heinz E (2000) Novel processive and nonprocessive glycosyltransferases from Staphylococcus aureus and Arabidopsis thaliana synthesize glycoglycerolipids, glycophospholipids, glycosphingolipids and glycosylsterols. European journal of biochemistry / FEBS 267: 3770-3783

Kanagarajan S, Muthusamy S, Gliszczynska A, Lundgren A, Brodelius PE (2012) Functional expression and characterization of sesquiterpene synthases from Artemisia annua L. using transient expression system in Nicotiana benthamiana. Plant Cell Rep 31: 1309-1319

Kannangara R, Motawia MS, Hansen NK, Paquette SM, Olsen CE, Moller BL, Jorgensen K (2011) Characterization and expression profile of two UDP-glucosyltransferases, UGT85K4 and UGT85K5, catalyzing the last step in cyanogenic glucoside biosynthesis in cassava. Plant J 68: 287-301

Keukens EA, de Vrije T, van den Boom C, de Waard P, Plasman HH, Thiel F, Chupin V, Jongen WM, de Kruijff B (1995) Molecular basis of glycoalkaloid induced membrane disruption. Biochim Biophys Acta 1240: 216-228

Kim BG, Kim HJ, Ahn JH (2012) Production of bioactive flavonol rhamnosides by expression of plant genes in Escherichia coli. Journal of agricultural and food chemistry 60: 11143-11148

Kim JH, Kim BG, Park Y, Ko JH, Lim CE, Lim J, Lim Y, Ahn JH (2006) Characterization of flavonoid 7-O-glucosyltransferase from Arabidopsis thaliana. Bioscience, biotechnology, and biochemistry 70: 1471-1477

Kitajima J, Kamoshita A, Ishikawa T, Takano A, Fukuda T, Isoda S, Ida Y (2003) Glycosides of Atractylodes lancea. Chemical & pharmaceutical bulletin 51: 673-678

Klee HJ (2013) Purple tomatoes: longer lasting, less disease, and better for you. Current biology : CB 23: R520-521

Ko JH, Kim BG, Hur HG, Lim Y, Ahn JH (2006) Molecular cloning, expression and characterization of a glycosyltransferase from rice. Plant Cell Rep 25: 741-746

Ko JH, Kim BG, Kim JH, Kim H, Lim CE, Lim J, Lee C, Lim Y, Ahn JH (2008) Four glucosyltransferases from rice: cDNA cloning, expression, and characterization. J Plant Physiol 165: 435-444

Kohara A, Nakajima C, Hashimoto K, Ikenaga T, Tanaka H, Shoyama Y, Yoshida S, Muranaka T (2005) A novel glucosyltransferase involved in steroid saponin biosynthesis in Solanum aculeatissimum. Plant Mol Biol 57: 225-239

Kohara A, Nakajima C, Yoshida S, Muranaka T (2007) Characterization and engineering of glycosyltransferases responsible for steroid saponin biosynthesis in Solanaceous plants. Phytochemistry 68: 478-486

Kong CH, Zhao H, Xu XH, Wang P, Gu Y (2007) Activity and allelopathy of soil of flavone o-glycosides from rice. Journal of agricultural and food chemistry 55: 6007-6012

Kubo A, Arai Y, Nagashima S, Yoshikawa T (2004) Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. Arch Biochem Biophys 429: 198-203

Kuete V, Efferth T (2013) Molecular determinants of cancer cell sensitivity and resistance towards the sesquiterpene farnesol. Die Pharmazie 68: 608-615

Kunii M, Kitahama Y, Fukushima EO, Seki H, Muranaka T, Yoshida Y, Aoyama Y (2012) beta-Amyrin oxidation by oat CYP51H10 expressed heterologously in yeast cells: the first example of CYP51-dependent metabolism other than the 14-demethylation of sterol precursors. Biological & pharmaceutical bulletin 35: 801-804

Lagarda MJ, Garcia-Llatas G, Farre R (2006) Analysis of phytosterols in foods. J Pharm Biomed Anal 41: 1486-1496

Lairson LL, Henrissat B, Davies GJ, Withers SG (2008) Glycosyltransferases: structures, functions, and mechanisms. Annual review of biochemistry 77: 521-555

Lairson LL, Withers SG (2004) Mechanistic analogies amongst carbohydrate modifying enzymes. Chem Commun (Camb): 2243-2248

Landmann C, Fink B, Schwab W (2007) FaGT2: a multifunctional enzyme from strawberry (Fragaria x ananassa) fruits involved in the metabolism of natural and xenobiotic compounds. Planta 226: 417-428

Lanot A, Hodge D, Jackson RG, George GL, Elias L, Lim EK, Vaistij FE, Bowles DJ (2006) The glucosyltransferase UGT72E2 is responsible for monolignol 4-O-glucoside production in Arabidopsis thaliana. Plant J 48: 286-295

Lee S (2007) Artemisinin, promising lead natural product for various drug developments. Mini reviews in medicinal chemistry 7: 411-422

Leiherer A, Mundlein A, Drexel H (2013) Phytochemicals and their impact on adipose tissue inflammation and diabetes. Vascular pharmacology 58: 3-20

Li L, Modolo LV, Escamilla-Trevino LL, Achnine L, Dixon RA, Wang X (2007) Crystal structure of Medicago truncatula UGT85H2--insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. J Mol Biol 370: 951-963

Li Y, Baldauf S, Lim EK, Bowles DJ (2001) Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana. J Biol Chem 276: 4338-4343

Lim EK, Ashford DA, Hou B, Jackson RG, Bowles DJ (2004) Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. Biotechnology and bioengineering 87: 623-631

Lim EK, Baldauf S, Li Y, Elias L, Worrall D, Spencer SP, Jackson RG, Taguchi G, Ross J, Bowles DJ (2003) Evolution of substrate recognition across a multigene family of glycosyltransferases in Arabidopsis. Glycobiology 13: 139-145

Lim EK, Bowles D (2004) Plant production systems for bioactive small molecules. Curr Opin Biotechnol 23: 271-277

Lim EK, Doucet CJ, Hou B, Jackson RG, Abrams SR, Bowles DJ (2005) Resolution of (+)-abscisic acid using an Arabidopsis glycosyltransferase. Tetrahedron-Asymmetr 16: 143-147

Lim EK, Doucet CJ, Li Y, Elias L, Worrall D, Spencer SP, Ross J, Bowles DJ (2002) The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. J Biol Chem 277: 586-592

Literakova P, Lochman J, Zdrahal Z, Prokop Z, Mikes V, Kasparovsky T (2010) Determination of capsidiol in tobacco cells culture by HPLC. Journal of chromatographic science 48: 436-440

Liu Z, Liu Y, Pu Z, Wang J, Zheng Y, Li Y, Wei Y (2013) Regulation, evolution, and functionality of flavonoids in cereal crops. Biotechnology letters 35: 1765-1780

Loutre C, Dixon DP, Brazier M, Slater M, Cole DJ, Edwards R (2003) Isolation of a glucosyltransferase from Arabidopsis thaliana active in the metabolism of the persistent pollutant 3,4-dichloroaniline. Plant J 34: 485-493

Luang S, Cho JI, Mahong B, Opassiri R, Akiyama T, Phasai K, Komvongsa J, Sasaki N, Hua YL, Matsuba Y, Ozeki Y, Jeon JS, Cairns JR (2013) Rice Os9BGlu31 is a transglucosidase with the capacity to equilibrate phenylpropanoid, flavonoid, and phytohormone glycoconjugates. J Biol Chem 288: 10111-10123

Macias FA, Marin D, Oliveros-Bastidas A, Chinchilla D, Simonet AM, Molinillo JM (2006) Isolation and synthesis of allelochemicals from gramineae: benzoxazinones and related compounds. Journal of agricultural and food chemistry 54: 991-1000

Mackenzie PI, Owens IS, Burchell B, Bock KW, Bairoch A, Belanger A, Fournel-Gigleux S, Green M, Hum DW, Iyanagi T, Lancet D, Louisot P, Magdalou J, Chowdhury JR, Ritter JK, Schachter H, Tephly TR, Tipton KF, Nebert DW (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. Pharmacogenetics 7: 255-269

Madhav H, Bhasker S, Chinnamma M (2013) Functional and structural variation of uridine diphosphate glycosyltransferase (UGT) gene of Stevia rebaudiana-UGTSr involved in the synthesis of rebaudioside A. Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale 63: 245-253

Maier W, Hammer K, Dammann U, Schulz B, Strack D (1997) Accumulation of sesquiterpenoid cyclohexenone derivatives induced by an arbuscular mycorrhizal fungus in members of the Poaceae. Planta 202: 36-42

Maier W, Peipp H, Schmidt J, Wray V, Strack D (1995) Levels of a terpenoid glycoside (blumenin) and cell wall-bound phenolics in some cereal mycorrhizas. Plant Physiol 109: 465-470

Martin RC, Mok MC, Habben JE, Mok DW (2001) A maize cytokinin gene encoding an O-glucosyltransferase specific to cis-zeatin. Proc Natl Acad Sci U S A 98: 5922-5926

Martin RC, Mok MC, Mok DW (1999) A gene encoding the cytokinin enzyme zeatin O-xylosyltransferase of Phaseolus vulgaris. Plant Physiol 120: 553-558

Masada S, Terasaka K, Oguchi Y, Okazaki S, Mizushima T, Mizukami H (2009) Functional and structural characterization of a flavonoid glucoside 1,6-glucosyltransferase from Catharanthus roseus. Plant & cell physiology 50: 1401-1415

Matros A, Mock HP (2004) Ectopic expression of a UDP-glucose:phenylpropanoid glucosyltransferase leads to increased resistance of transgenic tobacco plants against infection with Potato Virus Y. Plant & cell physiology 45: 1185-1193

Meech R, Miners JO, Lewis BC, Mackenzie PI (2012) The glycosidation of xenobiotics and endogenous compounds: versatility and redundancy in the UDP glycosyltransferase superfamily. Pharmacol Ther 134: 200-218

Meesapyodsuk D, Balsevich J, Reed DW, Covello PS (2007) Saponin biosynthesis in Saponaria vaccaria. cDNAs encoding beta-amyrin synthase and a triterpene carboxylic acid glucosyltransferase. Plant Physiol 143: 959-969

Messner B, Thulke O, Schaffner AR (2003) Arabidopsis glucosyltransferases with activities toward both endogenous and xenobiotic substrates. Planta 217: 138-146

Misawa N (2011) Pathway engineering for functional isoprenoids. Curr Opin Biotechnol 22: 627-633

Miyamoto K, Shimizu T, Lin F, Sainsbury F, Thuenemann E, Lomonossoff G, Nojiri H, Yamane H, Okada K (2012) Identification of an E-box motif responsible for the expression of jasmonic acid-induced chitinase gene OsChia4a in rice. J Plant Physiol 169: 621-627

Modolo LV, Li L, Pan H, Blount JW, Dixon RA, Wang X (2009) Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. J Mol Biol 392: 1292-1302

Moehs CP, Allen PV, Friedman M, Belknap WR (1997) Cloning and expression of solanidine UDP-glucose glucosyltransferase from potato. Plant J 11: 227-236

Mohamed AA, Ceunen S, Geuns JM, Van den Ende W, De Ley M (2011) UDP-dependent glycosyltransferases involved in the biosynthesis of steviol glycosides. J Plant Physiol 168: 1136-1141

Moheb A, Grondin M, Ibrahim RK, Roy R, Sarhan F (2013) Winter wheat hull (husk) is a valuable source for tricin, a potential selective cytotoxic agent. Food chemistry 138: 931-937

Moraga AR, Mozos AT, Ahrazem O, Gomez-Gomez L (2009) Cloning and characterization of a glucosyltransferase from Crocus sativus stigmas involved in flavonoid glucosylation. BMC Plant Biol 9: 109

Moraga AR, Nohales PF, Perez JA, Gomez-Gomez L (2004) Glucosylation of the saffron apocarotenoid crocetin by a glucosyltransferase isolated from Crocus sativus stigmas. Planta 219: 955-966

Moses T, Pollier J, Almagro L, Buyst D, Van Montagu M, Pedreno MA, Martins JC, Thevelein JM, Goossens A (2014) Combinatorial biosynthesis of sapogenins and saponins in Saccharomyces cerevisiae using a C-16alpha hydroxylase from Bupleurum falcatum. Proc Natl Acad Sci U S A 111: 1634-1639

Moses T, Pollier J, Thevelein JM, Goossens A (2013) Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro. New Phytol 200: 27-43

Mugford ST, Louveau T, Melton R, Qi X, Bakht S, Hill L, Tsurushima T, Honkanen S, Rosser SJ, Lomonossoff GP, Osbourn A (2013) Modularity of Plant Metabolic Gene Clusters: A Trio of Linked Genes That Are Collectively Required for Acylation of Triterpenes in Oat. Plant Cell 25: 1078–1092

Mugford ST, Milkowski C (2012) Serine carboxypeptidase-like acyltransferases from plants. Methods in enzymology 516: 279-297

Mugford ST, Qi X, Bakht S, Hill L, Wegel E, Hughes RK, Papadopoulou K, Melton R, Philo M, Sainsbury F, Lomonossoff GP, Roy AD, Goss RJ, Osbourn A (2009) A serine carboxypeptidase-like acyltransferase is required for synthesis of antimicrobial compounds and disease resistance in oats. Plant Cell 21: 2473-2484

Mylona P, Owatworakit A, Papadopoulou K, Jenner H, Qin B, Findlay K, Hill L, Qi X, Bakht S, Melton R, Osbourn A (2008) Sad3 and sad4 are required for saponin biosynthesis and root development in oat. Plant Cell 20: 201-212

Nagatoshi M, Terasaka K, Nagatsu A, Mizukami H (2011) Iridoid-specific glucosyltransferase from Gardenia jasminoides. J Biol Chem 286: 32866-32874

Nagatoshi M, Terasaka K, Owaki M, Sota M, Inukai T, Nagatsu A, Mizukami H (2012) UGT75L6 and UGT94E5 mediate sequential glucosylation of crocetin to crocin in Gardenia jasminoides. FEBS Lett 586: 1055-1061

Naoumkina MA, Modolo LV, Huhman DV, Urbanczyk-Wochniak E, Tang Y, Sumner LW, Dixon RA (2010) Genomic and coexpression analyses predict multiple

genes involved in triterpene saponin biosynthesis in Medicago truncatula. Plant Cell 22: 850-866

Nelson DR (2013) A world of cytochrome P450s. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 368: 20120430

Nishihara M, Nakatsuka T (2011) Genetic engineering of flavonoid pigments to modify flower color in floricultural plants. Biotechnology letters 33: 433-441

Nishikawa M, Nojima S, Akiyama T, Sankawa U, Inoue K (1984) Interaction of digitonin and its analogs with membrane cholesterol. Journal of biochemistry 96: 1231-1239

Noguchi A, Fukui Y, Iuchi-Okada A, Kakutani S, Satake H, Iwashita T, Nakao M, Umezawa T, Ono E (2008) Sequential glucosylation of a furofuran lignan, (+)-sesaminol, by Sesamum indicum UGT71A9 and UGT94D1 glucosyltransferases. Plant J 54: 415-427

Noguchi A, Horikawa M, Fukui Y, Fukuchi-Mizutani M, Iuchi-Okada A, Ishiguro M, Kiso Y, Nakayama T, Ono E (2009) Local differentiation of sugar donor specificity of flavonoid glycosyltransferase in Lamiales. Plant Cell 21: 1556-1572

Noguchi A, Saito A, Homma Y, Nakao M, Sasaki N, Nishino T, Takahashi S, Nakayama T (2007) A UDP-glucose:isoflavone 7-O-glucosyltransferase from the roots of soybean (glycine max) seedlings. Purification, gene cloning, phylogenetics, and an implication for an alternative strategy of enzyme catalysis. J Biol Chem 282: 23581-23590

Novy R, Drott D, Yaeger K, Mierendorf R (2001) Overcoming the codon bias of E. coli for enhanced protein expression. Innovations 12: 1-3

Nuccio ML, Rhodes D, McNeil SD, Hanson AD (1999) Metabolic engineering of plants for osmotic stress resistance. Curr Opin Plant Biol 2: 128-134

Nystrom L, Schar A, Lampi AM (2012) Steryl glycosides and acylated steryl glycosides in plant foods reflect unique sterol patterns. Eur. J. Lipid Sci. Technol 114: 656–669

Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ (2006) Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. EMBO J 25: 1396-1405

Ogata J, Kanno Y, Itoh Y, Tsugawa H, Suzuki M (2005) Plant biochemistry: anthocyanin biosynthesis in roses. Nature 435: 757-758

Oldroyd GE (2013) Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. Nature reviews. Microbiology 11: 252-263

Omura T (2013) Contribution of cytochrome P450 to the diversification of eukaryotic organisms. Biotechnology and applied biochemistry 60: 4-8

Ono E, Fukuchi-Mizutani M, Nakamura N, Fukui Y, Yonekura-Sakakibara K, Yamaguchi M, Nakayama T, Tanaka T, Kusumi T, Tanaka Y (2006) Yellow flowers generated by expression of the aurone biosynthetic pathway. Proc Natl Acad Sci U S A 103: 11075-11080

Ono E, Homma Y, Horikawa M, Kunikane-Doi S, Imai H, Takahashi S, Kawai Y, Ishiguro M, Fukui Y, Nakayama T (2010a) Functional differentiation of the

glycosyltransferases that contribute to the chemical diversity of bioactive flavonol glycosides in grapevines (Vitis vinifera). Plant Cell 22: 2856-2871

Ono E, Ruike M, Iwashita T, Nomoto K, Fukui Y (2010b) Co-pigmentation and flavonoid glycosyltransferases in blue Veronica persica flowers. Phytochemistry 71: 726-735

Osbourn A (1996) Saponins and plant defence - a soap story. Trends in plant science 1: 4-9

Osbourn A, Bowyer P, Lunness P, Clarke B, Daniels M (1995) Fungal pathogens of oat roots and tomato leaves employ closely related enzymes to detoxify different host plant saponins. Molecular plant-microbe interactions : MPMI 8: 971-978

Osmani SA, Bak S, Imberty A, Olsen CE, Moller BL (2008) Catalytic key amino acids and UDP-sugar donor specificity of a plant glucuronosyltransferase, UGT94B1: molecular modeling substantiated by site-specific mutagenesis and biochemical analyses. Plant Physiol 148: 1295-1308

Osmani SA, Bak S, Moller BL (2009) Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. Phytochemistry 70: 325-347

Ossola B, Kaariainen TM, Mannisto PT (2009) The multiple faces of quercetin in neuroprotection. Expert opinion on drug safety 8: 397-409

Owatworakit A, Townsend B, Louveau T, Jenner H, Rejzek M, Hughes RK, Saalbach G, Qi X, Bakht S, Deb Roy A, Mugford ST, Goss RJ, Field RA, Osbourn A (2012) Glycosyltransferases from oat (Avena) implicated in the acylation of avenacins. J Biol Chem 288: 3696-3704

Oyama T, Yasui Y, Sugie S, Koketsu M, Watanabe K, Tanaka T (2009) Dietary tricin suppresses inflammation-related colon carcinogenesis in male Crj: CD-1 mice. Cancer prevention research 2: 1031-1038

Pang Y, Peel GJ, Sharma SB, Tang Y, Dixon RA (2008) A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of Medicago truncatula. Proc Natl Acad Sci U S A 105: 14210-14215

Papadopoulou K, Melton RE, Leggett M, Daniels MJ, Osbourn AE (1999) Compromised disease resistance in saponin-deficient plants. Proc Natl Acad Sci U S A 96: 12923-12928

Peer WA, Murphy AS (2007) Flavonoids and auxin transport: modulators or regulators? Trends Plant Sci 12: 556-563

Peyret H, Lomonossoff GP (2013) The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. Plant Mol Biol 83: 51-58

Phillips DR, Rasbery JM, Bartel B, Matsuda SP (2006) Biosynthetic diversity in plant triterpene cyclization. Curr Opin Plant Biol 9: 305-314

Phillips K.M, Ruggio D.M, M. A-K (2004) ANALYSIS OF STERYL GLUCOSIDES IN FOODS AND DIETARY SUPPLEMENTS BY SOLID-PHASE EXTRACTION AND GAS CHROMATOGRAPHY. Journal of Food Lipids 12: 124–140

Piochon M, Legault J, Gauthier C, Pichette A (2009) Synthesis and cytotoxicity evaluation of natural alpha-bisabolol beta-D-fucopyranoside and analogues. Phytochemistry 70: 228-236

Piotrowska A, Bajguz A (2011) Conjugates of abscisic acid, brassinosteroids, ethylene, gibberellins, and jasmonates. Phytochemistry 72: 2097-2112

Poppenberger B, Fujioka S, Soeno K, George GL, Vaistij FE, Hiranuma S, Seto H, Takatsuto S, Adam G, Yoshida S, Bowles D (2005) The UGT73C5 of Arabidopsis thaliana glucosylates brassinosteroids. Proc Natl Acad Sci U S A 102: 15253-15258

Prasannan R, Kalesh KA, Shanmugam MK, Nachiyappan A, Ramachandran L, Nguyen AH, Kumar AP, Lakshmanan M, Ahn KS, Sethi G (2012) Key cell signaling pathways modulated by zerumbone: role in the prevention and treatment of cancer. Biochemical pharmacology 84: 1268-1276

Priest DM, Ambrose SJ, Vaistij FE, Elias L, Higgins GS, Ross AR, Abrams SR, Bowles DJ (2006) Use of the glucosyltransferase UGT71B6 to disturb abscisic acid homeostasis in Arabidopsis thaliana. Plant J 46: 492-502

Priest DM, Jackson RG, Ashford DA, Abrams SR, Bowles DJ (2005) The use of abscisic acid analogues to analyse the substrate selectivity of UGT71B6, a UDP-glycosyltransferase of Arabidopsis thaliana. FEBS Lett 579: 4454-4458

Qi X, Bakht S, Leggett M, Maxwell C, Melton R, Osbourn A (2004) A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. Proc Natl Acad Sci U S A 101: 8233-8238

Qi X, Bakht S, Qin B, Leggett M, Hemmings A, Mellon F, Eagles J, Werck-Reichhart D, Schaller H, Lesot A, Melton R, Osbourn A (2006) A different function for a member of an ancient and highly conserved cytochrome P450 family: from essential sterols to plant defense. Proc Natl Acad Sci U S A 103: 18848-18853

Ross J, Li Y, Lim E, Bowles DJ (2001) Higher plant glycosyltransferases. Genome Biol 2: REVIEWS3004

Sainsbury F, Lomonossoff GP (2008) Extremely high-level and rapid transient protein production in plants without the use of viral replication. Plant Physiol 148: 1212-1218

Sainsbury F, Thuenemann EC, Lomonossoff GP (2009) pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. Plant Biotechnol J 7: 682-693

Sawai S, Saito K (2011) Triterpenoid biosynthesis and engineering in plants. Front Plant Sci 2: 25

Sayama T, Ono E, Takagi K, Takada Y, Horikawa M, Nakamoto Y, Hirose A, Sasama H, Ohashi M, Hasegawa H, Terakawa T, Kikuchi A, Kato S, Tatsuzaki N, Tsukamoto C, Ishimoto M (2012) The Sg-1 glycosyltransferase locus regulates structural diversity of triterpenoid saponins of soybean. Plant Cell 24: 2123-2138

Schilmiller AL, Pichersky E, Last RL (2012) Taming the hydra of specialized metabolism: how systems biology and comparative approaches are revolutionizing plant biochemistry. Curr Opin Plant Biol 15: 338-344

Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K, Muranaka T (2008) Licorice beta-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. Proc Natl Acad Sci U S A 105: 14204-14209

Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from Medicago truncatula. Plant Cell 17: 3141-3154

Shibuya M, Hoshino M, Katsube Y, Hayashi H, Kushiro T, Ebizuka Y (2006) Identification of beta-amyrin and sophoradiol 24-hydroxylase by expressed sequence tag mining and functional expression assay. FEBS J 273: 948-959

Shibuya M, Nishimura K, Yasuyama N, Ebizuka Y (2010) Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in Glycine max. FEBS Lett 584: 2258-2264

Simons K, Sampaio JL (2011) Membrane organization and lipid rafts. Cold Spring Harbor perspectives in biology 3: a004697

Sonderby IE, Geu-Flores F, Halkier BA (2010) Biosynthesis of glucosinolates--gene discovery and beyond. Trends Plant Sci 15: 283-290

Soriano IR, Asenstorfer RE, Schmidt O, Riley IT (2004) Inducible Flavone in Oats (Avena sativa) Is a Novel Defense Against Plant-Parasitic Nematodes. Phytopathology 94: 1207-1214

Sozer O, Kis M, Gombos Z, Ughy B (2011) Proteins, glycerolipids and carotenoids in the functional photosystem II architecture. Frontiers in bioscience 16: 619-643

Sparg SG, Light ME, van Staden J (2004) Biological activities and distribution of plant saponins. J Ethnopharmacol 94: 219-243

Su ZH, Xu ZS, Peng RH, Tian YS, Zhao W, Han HJ, Yao QH, Wu AZ (2012) Phytoremediation of trichlorophenol by Phase II metabolism in transgenic Arabidopsis overexpressing a Populus glucosyltransferase. Environ Sci Technol 46: 4016-4024

Sui X, Gao X, Ao M, Wang Q, Yang D, Wang M, Fu Y, Wang L (2011) cDNA cloning and characterization of UDP-glucose: anthocyanidin 3-O-glucosyltransferase in Freesia hybrida. Plant Cell Rep 30: 1209-1218

Sun HX, Xie Y, Ye YP (2009) Advances in saponin-based adjuvants. Vaccine 27: 1787-1796

Suzuki H, Hayase H, Nakayama A, Yamaguchi I, Asami T, Nakajima M (2007) Identification and characterization of an Ipomoea nil glucosyltransferase which metabolizes some phytohormones. Biochemical and biophysical research communications 361: 980-986

Taguchi G, Yazawa T, Hayashida N, Okazaki M (2001) Molecular cloning and heterologous expression of novel glucosyltransferases from tobacco cultured cells that have broad substrate specificity and are induced by salicylic acid and auxin. European journal of biochemistry / FEBS 268: 4086-4094

Takos AM, Knudsen C, Lai D, Kannangara R, Mikkelsen L, Motawia MS, Olsen CE, Sato S, Tabata S, Jorgensen K, Moller BL, Rook F (2011) Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in Lotus japonicus and suggests the repeated evolution of this chemical defence pathway. Plant J 68: 273-286

Tanaka O, Tamura Y, Masuda H, Mizutani K (1996) Application of saponins in foods and cosmetics: saponins of Mohave yucca and Sapindus mukurossi. Advances in experimental medicine and biology 405: 1-11

Terpe K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Appl Microbiol Biotechnol 60: 523-533

Theis N, Raguso RA (2005) The effect of pollination on floral fragrance in thistles. Journal of chemical ecology 31: 2581-2600

Thimmappa R, Geisler K, Louveau T, O'Maille P, Osbourn A (2014) Triterpene Biosynthesis in Plants. Annu Rev Plant Biol

Threlfall DR, Whitehead IM (1988) Coordinated Inhibition of Squalene Synthetase and Induction of Enzymes of Sesquiterpenoid Phytoalexin Biosynthesis in Cultures of Nicotiana-Tabacum. Phytochemistry 27: 2567-2580

Tognetti VB, Van Aken O, Morreel K, Vandenbroucke K, van de Cotte B, De Clercq I, Chiwocha S, Fenske R, Prinsen E, Boerjan W, Genty B, Stubbs KA, Inze D, Van Breusegem F (2010) Perturbation of indole-3-butyric acid homeostasis by the UDP-glucosyltransferase UGT74E2 modulates Arabidopsis architecture and water stress tolerance. Plant Cell 22: 2660-2679

Tohge T, Kusano M, Fukushima A, Saito K, Fernie AR (2011) Transcriptional and metabolic programs following exposure of plants to UV-B irradiation. Plant signaling & behavior 6: 1987-1992

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

Treutter D (2005) Significance of flavonoids in plant resistance and enhancement of their biosynthesis. Plant biology 7: 581-591

Trossat C, Rathinasabapathi B, Weretilnyk EA, Shen TL, Huang ZH, Gage DA, Hanson AD (1998) Salinity promotes accumulation of 3-dimethylsulfoniopropionate and its precursor S-methylmethionine in chloroplasts. Plant Physiol 116: 165-171

Turlings TC, Tumlinson JH (1992) Systemic release of chemical signals by herbivore-injured corn. Proc Natl Acad Sci U S A 89: 8399-8402

Turlings TC, Tumlinson JH, Lewis WJ (1990) Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. Science 250: 1251-1253

Valero Galvan J, Valledor L, Gonzalez Fernandez R, Navarro Cerrillo RM, Jorrin-Novo JV (2012) Proteomic analysis of Holm oak (Quercus ilex subsp. ballota [Desf.] Samp.) pollen. Journal of proteomics 75: 2736-2744

van der Heide RF (1966) Reagents for the detection of antioxidants on thin layers of silica. Journal of chromatography 24: 239-243

van Herpen TW, Cankar K, Nogueira M, Bosch D, Bouwmeester HJ, Beekwilder J (2010) Nicotiana benthamiana as a production platform for artemisinin precursors. PloS one 5: e14222

Vetter J (2000) Plant cyanogenic glycosides. Toxicon : official journal of the International Society on Toxinology 38: 11-36

Vincken JP, Heng L, de Groot A, Gruppen H (2007) Saponins, classification and occurrence in the plant kingdom. Phytochemistry 68: 275-297

Vogt T, Jones P (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends Plant Sci 5: 380-386

von Rad U, Huttl R, Lottspeich F, Gierl A, Frey M (2001) Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize. Plant J 28: 633-642

Vrielink A, Ruger W, Driessen HP, Freemont PS (1994) Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose. EMBO J 13: 3413-3422

Wagner S, Klepsch MM, Schlegel S, Appel A, Draheim R, Tarry M, Hogbom M, van Wijk KJ, Slotboom DJ, Persson JO, de Gier JW (2008) Tuning Escherichia coli for membrane protein overexpression. Proc Natl Acad Sci U S A 105: 14371-14376

Wang J, Ma XM, Kojima M, Sakakibara H, Hou BK (2011) N-glucosyltransferase UGT76C2 is involved in cytokinin homeostasis and cytokinin response in Arabidopsis thaliana. Plant & cell physiology 52: 2200-2213

Wang J, Ma XM, Kojima M, Sakakibara H, Hou BK (2013a) Glucosyltransferase UGT76C1 finely modulates cytokinin responses via cytokinin N-glucosylation in Arabidopsis thaliana. Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale 65: 9-16

Wang KC, Ohnuma S (2000) Isoprenyl diphosphate synthases. Biochim Biophys Acta 1529: 33-48

Wang L, Han W, Xie C, Hou J, Fang Q, Gu J, Wang PG, Cheng J (2013b) Comparing the acceptor promiscuity of a Rosa hybrida glucosyltransferase RhGT1 and an engineered microbial glucosyltransferase OleD(PSA) toward a small flavonoid library. Carbohydr Res 368: 73-77

Wang X (2009) Structure, mechanism and engineering of plant natural product glycosyltransferases. FEBS Lett 583: 3303-3309

Wang YD, Zhang GJ, Qu J, Li YH, Jiang JD, Liu YB, Ma SG, Li Y, Lv HN, Yu SS (2013c) Diterpenoids and sesquiterpenoids from the roots of Illicium majus. Journal of natural products 76: 1976-1983

Warnecke DC, Baltrusch M, Buck F, Wolter FP, Heinz E (1997) UDP-glucose:sterol glucosyltransferase: cloning and functional expression in Escherichia coli. Plant Mol Biol 35: 597-603

Wegel E, Koumproglou R, Shaw P, Osbourn A (2009) Cell type-specific chromatin decondensation of a metabolic gene cluster in oats. Plant Cell 21: 3926-3936

Weis M, Lim EK, Bruce N, Bowles D (2006) Regioselective glucosylation of aromatic compounds: screening of a recombinant glycosyltransferase library to identify biocatalysts. Angew Chem Int Ed Engl 45: 3534-3538

Weston LA, Mathesius U (2013) Flavonoids: their structure, biosynthesis and role in the rhizosphere, including allelopathy. Journal of chemical ecology 39: 283-297

White JM, Shah P, Sanford S, Valverde M, Meier G (2010) Evaluation of Low Concentrations of Steryl Glucosides in Biodiesel by Gas Chromatography. Renewable Energy Group

Wink M (2003) Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. Phytochemistry 64: 3-19

Wu J, Filutowicz M (1999) Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. Acta biochimica Polonica 46: 591-599

Wu S, Jiang Z, Kempinski C, Eric Nybo S, Husodo S, Williams R, Chappell J (2012) Engineering triterpene metabolism in tobacco. Planta 236: 867-877

Xu R, Fazio GC, Matsuda SP (2004) On the origins of triterpenoid skeletal diversity. Phytochemistry 65: 261-291

Xu ZJ, Nakajima M, Suzuki Y, Yamaguchi I (2002) Cloning and characterization of the abscisic acid-specific glucosyltransferase gene from adzuki bean seedlings. Plant Physiol 129: 1285-1295

Xue Z, Duan L, Liu D, Guo J, Ge S, Dicks J, P OM, Osbourn A, Qi X (2012) Divergent evolution of oxidosqualene cyclases in plants. New Phytol 193: 1022-1038

Yamazaki M, Gong Z, Fukuchi-Mizutani M, Fukui Y, Tanaka Y, Kusumi T, Saito K (1999) Molecular cloning and biochemical characterization of a novel anthocyanin 5-O-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin. J Biol Chem 274: 7405-7411

Yang DH, Baldwin IT, Wu J (2013) Silencing brassinosteroid receptor BRI1 impairs herbivory-elicited accumulation of jasmonic acid-isoleucine and diterpene glycosides, but not jasmonic acid and trypsin proteinase inhibitors in Nicotiana attenuata. Journal of integrative plant biology 55: 514-526

Yang M, Sun J, Lu Z, Chen G, Guan S, Liu X, Jiang B, Ye M, Guo DA (2009) Phytochemical analysis of traditional Chinese medicine using liquid chromatography coupled with mass spectrometry. Journal of chromatography. A 1216: 2045-2062

Yesilada E, Bedir E, Calis I, Takaishi Y, Ohmoto Y (2005) Effects of triterpene saponins from Astragalus species on in vitro cytokine release. J Ethnopharmacol 96: 71-77

Yin R, Messner B, Faus-Kessler T, Hoffmann T, Schwab W, Hajirezaei MR, von Saint Paul V, Heller W, Schaffner AR (2012) Feedback inhibition of the general phenylpropanoid and flavonol biosynthetic pathways upon a compromised flavonol-3-O-glycosylation. Journal of experimental botany 63: 2465-2478

Yonekura-Sakakibara K, Fukushima A, Nakabayashi R, Hanada K, Matsuda F, Sugawara S, Inoue E, Kuromori T, Ito T, Shinozaki K, Wangwattana B, Yamazaki M, Saito K (2012) Two glycosyltransferases involved in anthocyanin modification delineated by transcriptome independent component analysis in Arabidopsis thaliana. Plant J 69: 154-167

Yonekura-Sakakibara K, Hanada K (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. Plant J 66: 182-193

Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis. Plant Cell 20: 2160-2176

Yonekura-Sakakibara K, Tohge T, Niida R, Saito K (2007) Identification of a flavonol 7-O-rhamnosyltransferase gene determining flavonoid pattern in Arabidopsis

by transcriptome coexpression analysis and reverse genetics. J Biol Chem 282: 14932-14941

Yoon JA, Kim BG, Lee WJ, Lim Y, Chong Y, Ahn JH (2012) Production of a novel quercetin glycoside through metabolic engineering of Escherichia coli. Appl Environ Microbiol 78: 4256-4262

Yoshihara N, Imayama T, Fukuchi-Mizutani M, Okuhara H, Tanaka Y, Ino I, Yabuya T (2005) cDNA cloning and characterization of UDP-glucose: Anthocyanidin 3-O-glucosyltransferase in Iris hollandica. Plant Science 169: 496–501

Yu B, Sun J, Yang X (2012) Assembly of Naturally Occurring Glycosides, Evolved Tactics, and Glycosylation Methods. Acc Chem Res

Yu F, Utsumi R (2009) Diversity, regulation, and genetic manipulation of plant mono- and sesquiterpenoid biosynthesis. Cellular and molecular life sciences : CMLS 66: 3043-3052

Yu LM (1995) Elicitins from Phytophthora and basic resistance in tobacco. Proc Natl Acad Sci U S A 92: 4088-4094

Zhu JY, Sae-Seaw J, Wang ZY (2013) Brassinosteroid signalling. Development 140: 1615-1620

Zhuang PY, Zhang GJ, Wang XJ, Zhang Y, Yu SS, Ma SG, Liu YB, Qu J, Li Y, Chen NH (2013) Novel Sesquiterpenoid Glycosides and Sesquiterpenes from the Roots of Illicium henryi. Planta medica 79: 1453-1460