Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume *Medicago truncatula*

Lahoucine Achnine, David V. Huhman, Mohamed A. Farag, Lloyd W. Sumner, Jack W. Blount and Richard A. Dixon[*]
Plant Biology Division, Samuel Roberts Noble Foundation, 2510 Sam Noble Parkway, Ardmore, OK 73401, USA

Summary

The biosynthesis of triterpene saponins is poorly characterized in spite of the importance of these glycosylated secondary metabolites for plant defense and animal health. The model legume *Medicago truncatula* synthesizes more than 30 different saponins based on at least five triterpene aglycones; soyasapogenols B and E, medicagenic acid, hederagenin and bayogenin. We have employed an inducible cell culture system, DNA array-based and *in silico* transcript profiling, and targeted metabolite profiling, to identify triterpene glycosyltransferases (GTs) from among the more than 300 GTs expressed in *M. truncatula*. Two uridine diphosphate glucosyltransferases were functionally characterized; UGT73K1 with specificity for hederagenin and soyasapogenols B and E, and UGT71G1 with specificity for medicagenic acid. The latter enzyme also glycosylated certain isoflavones and the flavonol quercetin with higher efficiency than triterpenes; however, integrated transcript and metabolite profiling supported a function for UGT71G1 in terpenoid but not (iso)flavonoid biosynthesis in the elicited cell cultures.

Keywords: triterpene, saponin, model legume, DNA arrays, *E. coli* expression.

Introduction

Triterpene saponins are a class of plant natural products with a wide range of bioactivities. Many display allelopathic, anti-microbial, or anti-insect repellent activity (Osbourn, 2003; Tava and Odoardi, 1996; Waller *et al.*, 1993). They can also be noxious to monogastric animals, reduce palatability of forages, or negatively impact forage digestibility in ruminants (Oleszek *et al.*, 1999; Small, 1996). However, saponins also have potentially useful pharmacological properties, including anticholesterolemic, anti-cancer, adjuvant, and hemolytic activities (Behboudi *et al.*, 1999; Haridas *et al.*, 2001; Jones and Elliott, 1969), and are important components of a number of herbal medicines (Chan *et al.*, 2002; Xu, 2001). In spite of the potential value of increasing or reducing levels of triterpene saponins for crop improvement, or their use as lead compounds for drug discovery, most of the genes involved in the biosynthesis of these complex molecules remain uncharacterized.

The saponins of the model legume *Medicago truncatula* are produced in roots and aerial plant parts, and are glycosides of at least five different triterpene aglycones; medicagenic acid, hederagenin, soyasapogenol B,

soyasapogenol E and bayogenin (Huhman and Sumner, 2002). These aglycones are derived from β-amyrin, a product of 2, 3-oxidosqualene formed by the terpene cyclase β-amyrin synthase (Kushiro *et al.*, 1998). The downstream reactions in the biosynthesis of *M. truncatula* saponins are believed to include a set of cytochrome P450-dependent hydroxylations/oxidations and several glycosyl transfer reactions catalyzed by glycosyltransferases (GTs) (Figure 1). To date, no plant GT involved in this pathway has been characterized at the molecular level, although a UDP-glucuronic acid:soyasapogenol glucuronosyltransferase has been characterized biochemically from germinating soybean seeds (Kurosawa *et al.*, 2002), UDP-glucose transferases active against sterol and steroidal glycoalkaloid aglycones have been cloned (Moehs *et al.*, 1997; Warnecke *et al.*, 1997), and a glycosyltransferase gene has been shown to be part of a cluster of genes in the oat genome involved in avenacin biosynthesis (Qi *et al.*, 2004).

Transcript profiling, when combined with genetic resources such as T-DNA insertion or activation-tagged mutants, or with elicitor-inducible cell systems, provides a
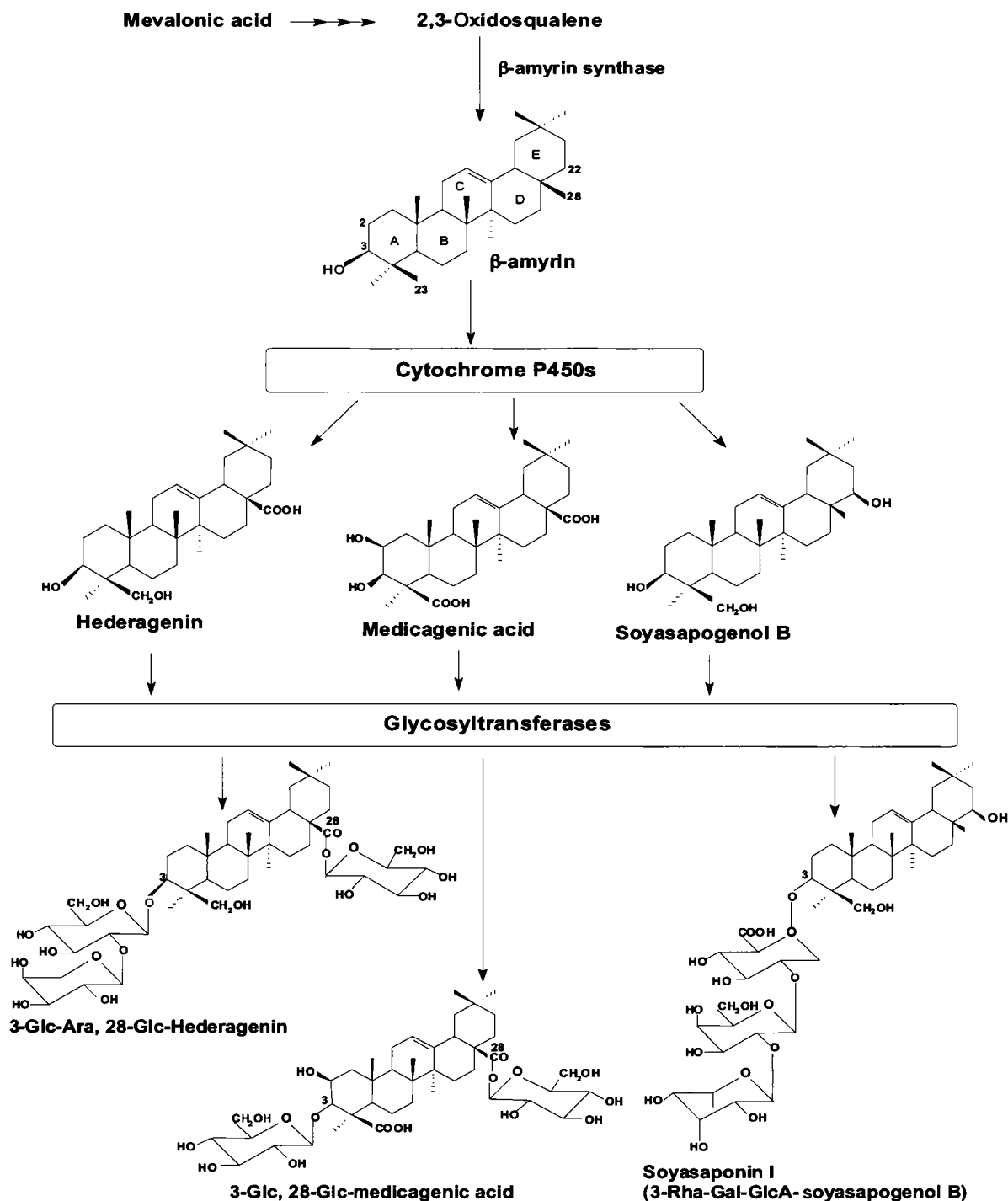
Figure 1. Diagrammatic representation of the saponin biosynthetic pathway in *Medicago truncatula*. The universal precursor β-amyrin is converted by a series of oxidative reactions to at least five aglycones (three of which are shown), which are converted by glycosyltransferases to over 35 different triterpene saponins.

powerful tool to accelerate gene discovery in natural product pathways (Dixon, 2001; Goossens *et al.*, 2003). Genetic and genomic approaches have recently been applied to triterpene biosynthesis. Classical mutagenesis is uncovering the avenacin pathway in oats (*Avena* spp.; Osbourn, 2003; Papadopoulou *et al.*, 1999), and expressed sequence tag

(EST) sequencing is being used as a gene discovery tool for ginsenoside biosynthesis in *Panax ginseng* (Jung *et al.*, 2003). In *M. truncatula* cell suspension cultures, exposure to methyl jasmonate (MeJA) results in accumulation of saponins preceded by induction of genes encoding three early enzymes of triterpene aglycone formation, squalene

synthase, squalene epoxidase, and β-amyrin synthase (β-AS; Suzuki *et al.*, 2002, 2005). By coupling DNA array approaches to profile transcripts in MeJA-induced and control cell cultures with metabolite profiling and *in silico* EST data mining, it has proved possible to select, and functionally characterize, GTs active in the triterpene pathway from among the more than 300 different GTs encoded in the *Medicago* genome.
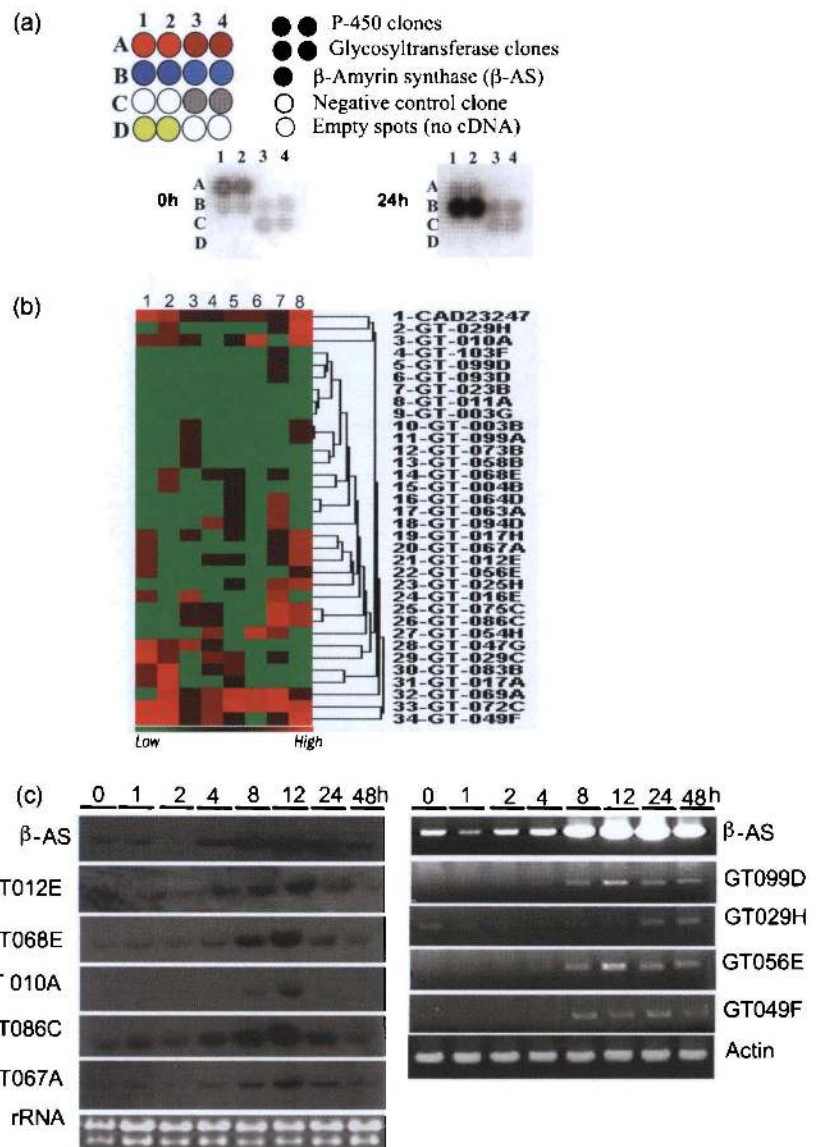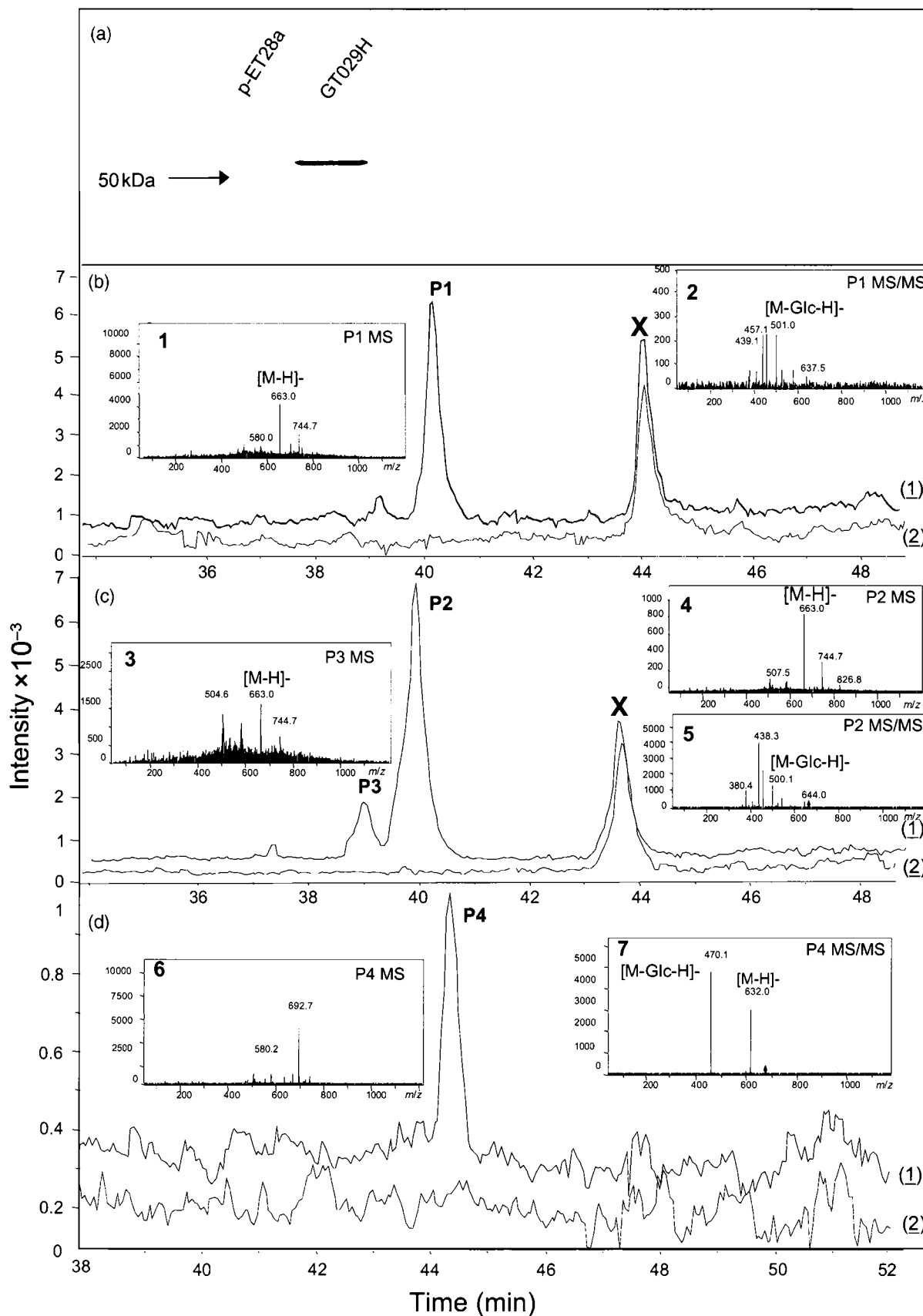
Results

Selection of target GTs

Mining all publicly available *M. truncatula* EST libraries (as of January 1, 2004) yielded 189 apparently distinct GTs (from multiple GT families) with two or more ESTs present in

tentative consensus (TC) sequences, and 114 GTs present as singletons. Family 1 UDP glycosyltransferases (UGTs) usually utilize small-molecular-weight compounds as acceptor substrates and UDP-sugars as donors (Li *et al.*, 2001), and commonly possess a carboxy terminal consensus sequence (the PSPG box) believed to be involved in binding to the UDP moiety of the sugar nucleotide donor (Vogt and Jones, 2000). Considering only these UGTs reduced the total number of candidate clones to 164, of which 63 existed in the Noble Foundation's EST collection as full-length cDNA clones.

Although some UGTs have recently been cloned and functionally characterized (Hansen *et al.*, 2003; Jones *et al.*, 2003; Kramer *et al.*, 2003; Lim *et al.*, 2002), it is not yet possible to use primary sequence as a predictor of function, beyond the presence of the PSPG box that distinguishes UGTs from the processive enzymes involved in polysaccharide

Figure 2. Selection of candidate triterpene GTs by DNA array and *in silico* expression analysis.
(a) Design of DNA macroarrays and autoradiograph showing portion of array with upregulated GT candidates. Note that one of the two GTs on the grid section is very strongly upregulated, the other by only threefold (in parallel with the level of β-AS transcript induction in this experiment).
(b) Cluster-histogram displaying the *in silico* expression of candidate *Medicago truncatula* MeJA-upregulated GTs and β-AS in the eight EST libraries indicated. The data set sizes (total number of ESTs sequenced in each library as of June 2004) for the individual libraries were: 1– MeJA-elicited cell culture (6900); 2 – nitrogen-starved root (7939); 3 – drought-stressed seedling (9520); 4 – mycorrhizal root (infected with *Glomus versiforme*) (7368); 5 – root, early stage (1–4 days) of nodulation (10 647); 6 – germinating seed (1524); 7 – developing stem (10 783); and 8 – leaf exposed to insect herbivory (10 309). The cluster histogram was generated using CLUSFAVOR software © 1991–2001, Leif E. Peterson, Version 4.0.
(c) Kinetics of induction of *M. truncatula* GTs and β-AS in response to MeJA, as determined by RNA gel blot or RT-PCR analysis. Left panel, RNA gel blot; right panel, RT-PCR analysis.

(a)

p-ET28a

GT029H

50 kDa →

(b)

7
6
5
4
3
2
1
0

P1

X

1 P1 MS

10000
8000
6000
4000
2000
0

[M-H]-
663.0

580.0   744.7

200   400   600   800   1000   *m/z*

2 P1 MS/MS

500
400
300
200
100
0

[M-Glc-H]-

457.1 501.0
439.1

637.5

200   400   600   800   1000   *m/z*

(1)

(2)

36   38   40   42   44   46   48

(c)

7
6
5
4
3
2
1
0

P2

P3

X

3 P3 MS

2500

1500

500
0

[M-H]-
663.0
504.6

744.7

200   400   600   800   1000   *m/z*

4 P2 MS

1000
800
600
400
200
0

[M-H]-
663.0

744.7
507.5
826.8

200   400   600   800   1000   *m/z*

5 P2 MS/MS

5000
4000
3000
2000
1000
0

438.3

[M-Glc-H]-

380.4
500.1
644.0

200   400   600   800   1000   *m/z*

(1)

(2)

36   38   40   42   44   46   48

(d)

1

0.8

0.6

0.4

0.2

0

P4

6 P4 MS

10000

7500

5000

2500

0

692.7

580.2

200   400   600   800   1000   *m/z*

7 P4 MS/MS

5000

4000

3000

2000

1000

0

470.1

[M-Glc-H]-

[M-H]-
632.0

200   400   600   800   1000   *m/z*

(1)

(2)

38   40   42   44   46   48   50   52

Time (min)

Intensity ×10⁻³

biosynthesis. A transcript profiling approach was therefore adopted to reduce the number of candidate GTs. The 164 cDNAs were spotted onto duplicate membranes for macroarray hybridization analysis using labeled RNA from 24 h MeJA treated or control cell cultures (Figure 2a). Transcripts corresponding to 33 of the candidate GTs were upregulated by threefold or more in response to MeJA. The cluster-histogram (Figure 2b) shows the relative expression level of these 33 GTs in EST libraries (information available at http://www.tigr.org/tdb/mtgi/) where *M. truncatula* β-AS (GenBank ID CAD23247) is expressed. To further reduce the number of candidates, an arbitrary cut-off was imposed to eliminate GT clones that had an expression frequency higher than 0.05 (i.e., five ESTs per 10 000 clones) in any EST library where β-AS was not expressed. Eighteen GT candidates passed this filter, being co-expressed with β-AS in more than two cDNA libraries (Figure 2b).

Tight co-induction of genes encoding enzymes in a particular biosynthetic pathway is well documented (Batz *et al.*, 1998; Dixon *et al.*, 1995). RNA gel blot and RT-PCR analysis was therefore performed to compare the induction kinetics of the candidate GT transcripts with those of β-AS following exposure of *M. truncatula* cells to MeJA. Nine candidate GTs were co-induced with β-AS (Figure 2c), although their induction kinetics varied. Induction of GT010A was quite transient, with transcripts only detected at 8, 12 and 24 h (weak). Induction of GT012E, GT068E, GT086C, GT067A, GT099D, GT056E and GT049F more closely reflected that of β-AS, whereas GT029H was induced somewhat later than β-AS.

Functional characterization and substrate specificity of triterpene GTs

The nine candidate GTs were expressed in *Escherichia coli* BL21 cells. Only five of the enzymes (GT029H, GT049F, GT010A, GT086C and GT067A) were recovered in the soluble fraction. In the case of GT029H, SDS-PAGE analysis of total proteins showed that a 52-kDa band, corresponding to the predicted size (465 amino acids) of the recombinant protein, appeared in extracts from IPTG induced *E. coli*, but not in cultures harboring empty pET-28a vector (data not shown). Similar results were obtained with recombinant GT049F. The recombinant 52 kDa GT029H was His-tag purified (Figure 3a) and assayed against an array of acceptors including

triterpene aglycones and phenylpropanoids. UDP-glucose and UDP-galactose were tested as sugar donors; no products were obtained with the latter.
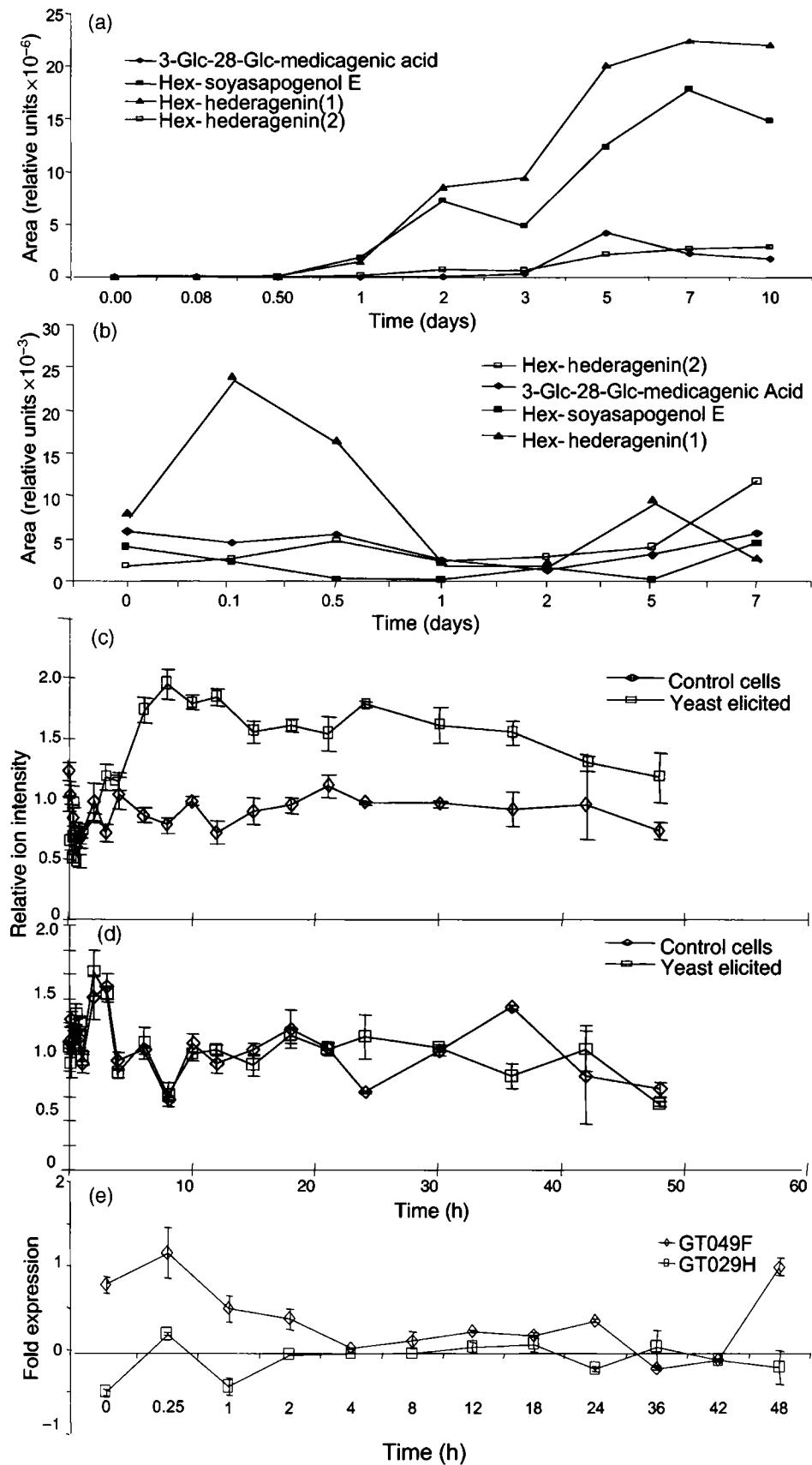
The products of GT029H using UDP-glucose as donor and *M. truncatula* (Figure 3b) or *M. sativa* (Figure 3c) triterpene aglycone mixtures as acceptors were analyzed by mass spectrometry. Both the *M. truncatula* and *M. sativa* triterpene preparations contained the same major aglycones medicagenic acid, hederagenin, soyasapogenol B, soyasapogenol E and bayogenin. Negative ion/LC/ESI/ITMS extracted-ion chromatograms showed products with retention times of 40.1 (P1) and 39.7 min (P2), which were identified as medicagenic acid mono-glucosides by MS and MS-MS (Figure 3b,c). Neither P1 nor P2 was medicagenic acid 3-glucoside, an authentic sample of which co-chromatographed with the peak marked X in Figure 3(b,c). GT029H was also active with pure hederagenin as substrate (Figure 3d) yielding a hederagenin mono-glucoside (P4) with retention time of 44.4 min, the basic structure of which was confirmed by MS-MS (Figure 3d). Soyasapogenol A and soyasapogenol B were not glycosylated by GT029H (data not shown).

Using the same expression, purification and analytical protocols, GT049F was shown to glycosylate soyasapogenol E, soyasapogenol B and hederagenin to single products (P5, P6 and P7, respectively; Figure S1).

The potential activity of GTs 029H and 049F against 32 different phenolic acceptors was tested by incubation of the purified recombinant enzymes using UDP-glucose, UDP-galactose or UDP-glucuronic acid and mixtures of the compounds (four groups with eight compounds per group). Compounds within each group are readily separated by the HPLC method, and each compound has a distinct UV spectrum that is unlike that of any other compound in that particular group. Group 1 consisted of catechin, 4-coumaric acid, scopoletin, dihydroquercetin, pelargonidin, daidzein, luteolin and genistein. Of these, only the isoflavone genistein was glucosylated by recombinant GT029H (Figure S2a). Group 2 contained caffeic acid, coniferyl alcohol, benzoic acid, dihydrokaempferol, quercetin, coumestrol, medicarpin and biochanin A. Of these compounds, only quercetin (a flavonol) and biochanin A (4'-O-methyl genistein) were substrates for GT029H, with efficient production of conjugates of retention time 38.5 and 45.5 min, respectively (Figure S2b,c). None of the remaining compounds in

---

Table 1 Properties of *Medicago truncatula* GT029H and GT049F

Enzyme	Substrate	Product	K_m (μm)	K_{cat} (sec^{-1})	K_{cat} K_m^{-1} (sec^{-1} M^{-1})
GT029H	Quercetin	Multiple products	25	0.12	4800
	Genistein	Genistein 7-O-glucoside	32	0.13	4062
	Biochanin A	Biochanin A 7-O-glucoside	45	0.19	4222
	Hederagenin	Hederagenin 3- or 28-O-glucoside	166	0.025	153
GT049F	Soyasapogenol B	Soyasapogenol B 3-, 22-, or 23-O-glucoside	165	0.14	848
	Hederagenin	Hederagenin 3- or 28-O-glucoside	235	0.12	511

groups 3 and 4 was a substrate for GT029H: 4-hydroxybenzoic acid, vanillin, ferulic acid, cyanidin, cinnamic acid, naringenin, kaempferol, formononetin, esculetin, epicatechin, delphinidin, salicylic acid, liquiritigenin, apigenin, maackiain, or prunetin.

A radioactive assay using ^{14}C-UDP-glucose was used to determine the kinetic properties of purified recombinant GT029H and GT049F for triterpene and phenolic substrates (Table 1). Overall, the catalytic efficiencies of the two enzymes were significantly lower than that shown for UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor* (K_{cat} K_m^{-1} 12 600 sec^{-1} M^{-1}; Hansen *et al.*, 2003). GT029H exhibited Michaelis–Menten kinetics with K_m values ranging from 166 μM for hederagenin to 25 μM for quercetin at a fixed UDPG concentration of 42 μM.

Integrated transcript and metabolite profiling to assess GT function in M. truncatula *cell cultures in vivo*

The K_{cat} K_m^{-1} values of GT029H for (iso)flavonoid substrates were 1 order of magnitude higher than for hederagenin, suggesting that quercetin, genistein or biochanin A, if present in *M. truncatula*, would serve as preferred substrates for GT029H. To address preferred substrates for GT029H and GT049F *in vivo*, we compared changes in GT transcript levels (determined by RT-PCR or DNA microarray analysis using a 16-000 oligonucleotide array (Suzuki *et al.*, 2005) with metabolite levels in cell cultures responding to either MeJA or yeast elicitor, a treatment that induces isoflavonoid but not triterpene pathway genes (Suzuki *et al.*, 2005). MeJA treatment resulted in induction of both GT029H and GT049F transcripts, as observed by RT-PCR (Figure 2c). Analysis of the same cell samples indicated that MeJA induced accumulation of triterpene glycosides (Figure 4a) but not flavonoid/isoflavonoid glycosides (data not shown). In contrast, an isoflavone glycoside (the 7-O-glucoside of genistein) was induced by YE (Figure 4c). The corresponding glucoside of

biochanin A was detected in the cultures but was not induced by YE (Figure 4d). DNA microarray analysis indicated that neither GT029F nor GT049F was induced by YE (Figure 4e). Furthermore, glycosides of quercetin, the preferred *in vitro* substrate of GT029H, were not detected in control or elicited *M. truncatula* cultures.

Over 30 different saponins were detected by LC-MS analysis of extracts from MeJA induced cells (Suzuki *et al.*, 2005); the full structural characterization of these compounds [i.e., the position(s) of glycosylation or the exact nature of hexose substituents] has yet to be achieved. The characterized compounds accumulating to the highest levels are rhamnose-hexose-hexose-soyasapogenol E, hexose-hexose-soyasapogenol E, hexose-soyasapogenol E, rhamnose-hexose-hexose-hederagenin, hexose-hederagenin, hexose-hexose-bayogenin, hexose-bayogenin, 3-Glc-28-Glc-medicagenic acid, and 3-Glc medicagenic acid (Suzuki *et al.*, 2005). No triterpenes were strongly induced before 24 h post-elicitation. Triterpenes were detected in YE-treated cultures (Figure 4b), but the levels were 3 orders of magnitude lower than in MeJA-treated cultures. Overall, appearance of the medicagenic acid conjugates in MeJA-treated cultures was delayed compared with the appearance of soyasapogenol E and hederagenin conjugates, and lower levels were attained (Figure 4a). These kinetics are consistent with the involvement of GT029H and GT049F in triterpene metabolism, and the specific substrate preferences of the two enzymes for the different triterpene aglycones (compare Figures 4a and 2c).

Phylogeny of M. truncatula *GTs*

Phylogenetic analysis of the two functionally characterized *Medicago* triterpene GTs and other functionally characterized members of the plant UGT superfamily indicated that the two triterpene UGTs belong to different GT families (Figure 5). Protein sequence alignment using the Clustal

---

Figure 4. Induction of GTs, triterpenes and isoflavone derivatives in elicited *Medicago truncatula* cell suspension cultures.
(a) Accumulation of 3-Glc-28-Glc-medicagenic acid, hexose-soyasapogenol E and hexose-hederagenin (1) and (2) (isomers with different positions of glycosylation) in response to MeJA.
(b) Accumulation of 3-Glc-28-Glc-medicagenic acid, hexose-soyasapogenol E, and hexose-hederagenin (1) and (2) in response to yeast elicitor.
(c) Accumulation of genistein 7-O-glucoside in response to yeast elicitor.
(d) Accumulation of biochanin-A-7-O-glucoside in response to yeast elicitor.
(e) Changes in GT049F and GT029H transcript levels (determined by DNA microarray analysis) in response to yeast elicitor.
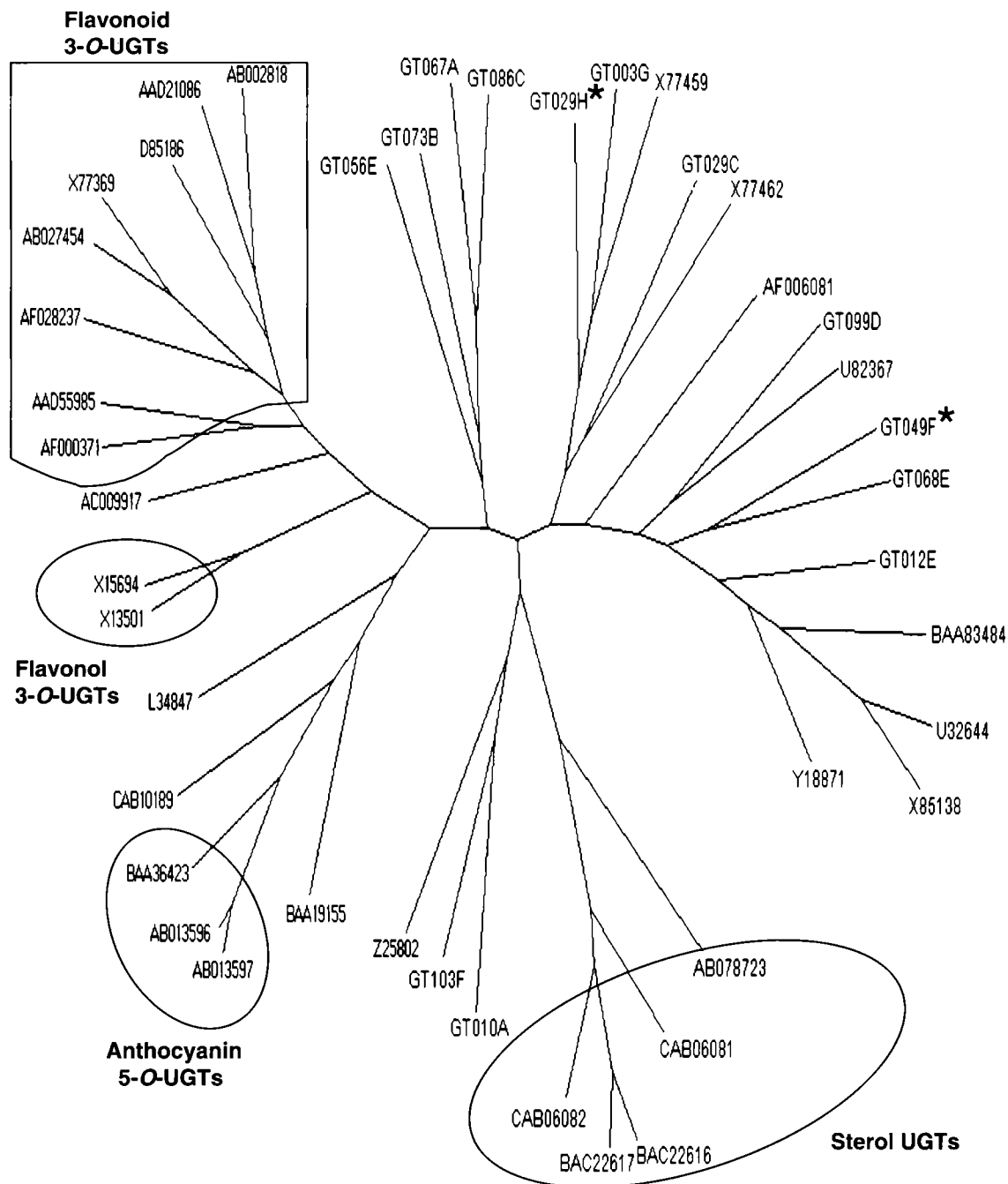
Figure 5. Unrooted tree dendrogram comparison of the amino acid sequences of triterpene glucosyltransferases (UGT71G1 and UGT73K1, labeled with asterisks), and other candidate *Medicago truncatula* UGTs (GT103F, GT010A, GT012E, GT068E, GT099D, GT029C, GT003G, GT086C, GT067A, GT073B, GT056E), with other functionally characterized plant glycosyltransferases.

The genes are: flavonoid-3-O-GTs; *Perilla frutescens* flavonoids 3-O-UGT GB ID AB002818, *Forsythia* × *intermedia* flavonoid 3-O-UGT GB ID AAD21086, *Gentiana triflora* flavonoid-3-*O*-UGT GB ID D85186, *Solanum melongena* flavonoid-3-*O*-UGT GB ID X77369, *Petunia* × *hybrida* anthocyanin-3-*O*-UGT GB ID AB027454, *Ipomoea purpurea* flavonoid 3-O-UGT GB ID AF028237, *Petunia* × *hybrida* UDP-Gal: flavonoid-3-O-UGT GB ID AAD55985, *Vitis vinifera* flavonoid 3-O-UGT GB ID AF000371: flavonol-3-*O*GTs; *Hordeum vulgare* flavonol 3-*O*-UGT GB ID X15694, *Zea mays* flavonol 3–*O*-UGT GB ID X13501: sterol UGTs; *Panax ginseng* sterol UGT GB ID BAC22617, *Avena sativa* sterol UGT GB ID CAB06081, *Aspergillus oryzae* sterol UGT GB ID AB078723, *Arabidopsis thaliana* sterol UGT GB ID CAB06082, *Panax ginseng* sterol 3-*O*-UGT GB ID BAC22616: anthocyanin-5-*O*-UGTs; Beefsteak-mint *Perilla frutescens* anthocyanin-5-*O*-UGTs GB ID AB013596 and AB013597, *Verbena* × *hybrida* anthocyanin 5-O-UGT GB ID BAA36423: other known plant UGTs; *Arabidopsis thaliana* UGT GB ID AC009917, *Dorotheantus bellidiformis* betanidin 5-*O*-UGT GB ID Y18871, *Lycopersicon esculentum* UGT GB ID X85138, *Scutellaria baicalensis* UDP-flavonoid 7-*O*-UGT GB ID BAA83484, *Solanum berthanltii* UGT GB ID AF006081, *Petunia hybrida* UDP-Rha: anthocyanidin 3-glucoside rhamnosyltransferase GB ID Z25802, *Nicotiana tabacum* jasmonate-induced UGT GB ID BAA19155, *Arabidopsis thaliana* UGT GB ID CAB10189, *Zea mays* indole 3-acetic acid 3-*O*-UGT GB ID L34847, *Nicotiana tabacum* salicylate-induced UGT GB ID U32644, *Manihot esculenta* Crantz UGTs, GB ID X77462 and X77459, *Solanum tuberosum* solanidine UGT, GB ID U82367.

program indicated that *M. truncatula* GT029H is an ortholog of *Arabidopsis thaliana* UGT71D2 (Li *et al.*, 2001), a putative GT yet to be functionally characterized, and GT029H has been assigned the designation UGT71G1. It is also closely related to cassava (*Manihot esculenta*) UGT (GenBank accession no. X77459) and *Solanum tuberosum* solanidine UGT (U82367) but not to *Avena sativa* sterol UGT (CAB06081) or *Panax ginseng* sterol UGT (BAC22617). Among other functionally characterized UGTs, *Forsythia × intermedia* flavonoid 3-*O*-UGT (AAD21086), *Petunia × hybrida* anthocyanin-3-*O*-UGT (AB027454) and *P. hybrida* UDP-Gal: flavonoid-3-*O*-UGT (AAD55985) form a separate cluster distinct from UGT71G1. GT049F represents the first member of a new UGT subfamily, and has been assigned the designation UGT73K1.

Tissue-specific expression and subcellular localization of triterpene GTs in M. truncatula

Expression patterns were determined by *in silico* mining of EST databases and by RT-PCR analysis of total RNA extracted from flower, leaf, root, stem and petiole of mature *M. truncatula* plants. UGT71G1 transcripts were only represented in EST libraries from leaves exposed to insect herbivory, nitrogen-starved roots and developing stem. The expression pattern of UGT73K1 was broader, with ESTs represented in seven of the eight libraries in which β-AS transcripts were found (Figure 2b). RT-PCR revealed expression of UGT71G1, UGT73K1 and β-AS in all five tissue types analyzed (Figure 6a). The levels of transcripts encoding the two GTs were almost similar in the different tissues, except that UGT71G1 transcripts were relatively higher in flower tissue. Consistent
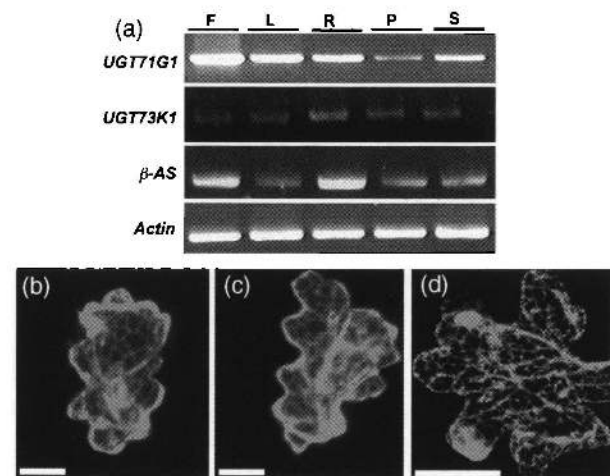
Figure 6. Expression patterns of triterpene UGTs.
(a) Tissue-specific expression using RT-PCR analysis of triterpene GTs and β-AS transcript levels in different tissues of *Medicago truncatula*. F, flowers; L, leaf, R, root, P, petiole, S, stem.
Subcellular localization of sGFP (b), GT029H-sGFP fusion (c) and sGFP-HDEL (d) following bombardment of expression constructs into *M. truncatula* leaf epidermal cells. Bars = 20 μm.

with this, a comparison of saponin profiles indicated that leaves had a higher relative abundance of medicagenic acid conjugates, whereas the non-acidic soyasapogenol conjugates were observed at higher relative abundance in roots (D.V. Huhman, M. Berhow, W. Olesezk, and L.W. Sumner, unpublished data). Interestingly, UGT71G1 is expressed at higher levels in the *M. truncatula* EST library from *Phoma medicaginis* infected leaves compared with non-infected leaf libraries, consistent with the well documented involvement of saponins as anti-fungal agents (Papadopoulou *et al.*, 1999; Zimmer *et al.*, 1967).

Particle bombardment of a UGT71G1-sGFP fusion into *M. truncatula* or *M. sativa* leaves revealed diffuse fluorescence, some in cytoplasmic strands. The pattern contrasted with that observed for sGFP-HDEL, the endoplasmic reticulum (ER) retention signal of which resulted in a characteristic reticulate pattern indicating typical ER localization (Figure 6b-d). Thus, although triterpene aglycones are hydrophobic molecules, and the pathway involves several membrane-bound enzymes (including a number of cytochrome P450s), it would appear that at least some steps of triterpene glycosylation take place in the cytoplasm.

Discussion

Genomics approaches for selecting target genes involved in secondary metabolism

A genomics approach using EST profiling, DNA macroarray analysis and metabolite profiling has led to the characterization of UGT71G1 and UGT73K1, MeJA-induced GTs from *M. truncatula* with activity for medicagenic acid/hederagenin and soyasapogenols B and E/hederagenin, respectively. Of the other GTs selected by this approach one (GT068E, expressed in soluble form only as a maltose-binding protein fusion) was subsequently found to be active with hederagenin as substrate in preliminary assays. The other soluble GTs (GT010A, GT086C and GT067A) were inactive against mixed or single triterpene aglycone preparations, and may possibly function to add additional sugars to a mono- or di-glycosidic saponin.

The approach taken shares similarities with coupled cDNA-AFLP or DNA microarray studies linked to metabolite profiling for gene discovery in alkaloid biosynthesis in tobacco (Goossens *et al.*, 2003) or glucosinolate biosynthesis in Arabidopsis (Hirai *et al.*, 2004), respectively. Saponins are produced constitutively in *Medicago* species, with higher levels in roots than in aerial parts (Nowacka and Oleszek, 1994), and the databases searched contained over 68 000 root ESTs, and 40 000 leaf/stem ESTs, suggesting a comprehensive coverage of potential triterpene GT sequences. However, it is likely that further GTs are involved in the formation of the more than 35 triterpene saponins of *M. truncatula*. Studies are in progress to functionally characterize the remaining

candidates from the present approach, and to globally profile GT expression at multiple times post-elicitation using DNA microarray analysis to ensure that potential candidates were not discarded by the imposition of the arbitrary cut-offs. Although saponin GT activity is inducible by MeJA (Hayashi *et al.*, 2003), it is also possible that one or more triterpene pathway GTs may be constitutively expressed and therefore missed by the present approach.

Properties of triterpene glycosyltransferases

Both UGT71G1 and UGT73K1 utilize UDP-glucose as opposed to UDP-galactose as sugar donor. This is consistent with the recent demonstration that the last amino acid of the PSPG motif controls relative specificity for UDP-glucose or UDP-galactose (Kubo *et al.*, 2004), this being a glutamine in glucosyltransferases (as in UGT71G1 and UGT73K1) and a histidine in galactosyltransferases. By this criterion, all nine candidate triterpene GTs from the macroarray analysis would be predicted to utilize UDP-glucose.

The regiospecificities of the two UGTs remain to be rigorously determined; this will depend upon the generation of sufficient product for NMR analysis. A UGT has been characterized that catalyzes transfer of glucose to the hydroxyl group of C-3 of ring A of the steroidal glycoalkaloid solanidine (Moehs *et al.*, 1997), a compound with similar overall structure to the triterpene aglycones. However, it is not closely related phylogenetically to UGT71G1. Furthermore, the C-3 hydroxyl of medicagenic acid is not targeted by UGT71G1, as a small amount of 3-Glc-medicagenic acid remained in the aglycone mixture used in the GT assays, and the level of this compound did not increase following incubation of the crude mixture, containing free medicagenic acid, with UGT71G1 and UDPG (Figure 3). It is therefore more likely that the available C-23 and C-28 positions are substituted, with formation of the glucose esters.

Several UGTs display a general regiospecificity rather than a tight substrate specificity (Hansen *et al.*, 2003; Jones *et al.*, 2003; Kramer *et al.*, 2003; Lim *et al.*, 2002, 2003), consistent with the present observations with UGT71G1. We have recently obtained crystals of UGT71G1 in complex with UDPG that diffract to 2.4 Å (X. Wang, L. Achnine, X.-Z. He and R.A. Dixon, unpublished data). Elucidation of the crystal structures of a range of plant UGTs will be necessary before the rules governing the substrate specificities of these versatile enzymes can be determined.

Previous studies using cellular fractionation or immunolocalization have suggested cytosolic (Anhalt and Weissenböck, 1992; Blume *et al.*, 1979; Schmid *et al.*, 1982; Yazaki *et al.*, 1995), vacuolar (Anhalt and Weissenböck, 1992), or vesicular/endomembrane (Latchinian-Sadek and Ibrahim, 1991) localizations for various plant natural product GTs. It has been suggested that GTs may be weakly associated with the endomembrane (Ibrahim, 1992) or with ER-localized

cytochrome P450s (Jones and Vogt, 2001). Based on the results of GFP-fusion protein localization, UGT71G1 appears to be a soluble cytoplasmic enzyme, at least when expressed in alfalfa leaf epidermal cells. In spite of the fact that the preceding enzymes in the triterpene pathway are likely to be membrane-bound cytochrome P450s, the localization pattern of UGT71G1-sGFP is clearly different for that observed for other operationally soluble enzymes of natural product biosynthesis that may associate with cytochrome P450 enzymes *in vivo* (Achnine *et al.*, 2004; Liu and Dixon, 2001).

In vitro *activity versus* in vivo *function for natural product pathway enzymes*

On the basis of *in vitro* kinetic measurements, UGT71G1 appears to have a clear preference for quercetin and 5-hydroxyisoflavones when compared with triterpenes. However, comparative transcript and metabolite profiling conclusively demonstrated that, at least in the elicited cell cultures studied in the present work, this enzyme is unlikely to be involved in (iso)flavonoid or flavonoid (quercetin glycoside) biosynthesis. Furthermore, UGT71G1 lacks activity with the 5-deoxy-isoflavonoid derivatives (formononetin and medicarpin) that occur as common glycosides in the roots of *Medicago* species. When the isoflavone genistein is produced ectopically in alfalfa leaves as a result of transformation with *M. truncatula* isoflavone synthase, it is converted to the 7-*O*-glucoside (by the action of UGT71G1 and/or other GTs?) whereas the 'natural' leaf flavonoids, such as apigenin, exist primarily as glucuronic acid conjugates (B. Deavours and R.A. Dixon, unpublished data). These observations raise an important problem in the interpretation of *in vitro* biochemical data, namely that many enzymes of natural product biosynthesis are promiscuous, and the most critical factor in determining their *in vivo* activity may be the availability of a suitable substrate. Similar arguments have been proposed in relation to the specificity of *O*-methylation reactions in the synthesis of flavor compounds in strawberry (Wein *et al.*, 2002) and lignin in alfalfa (Dixon *et al.*, 2001). Because of this problem, neither *in vitro* functional determination, nor, in the case of multigene families such as UGTs with obvious functional redundancy, simple genetic knock-out approaches, will provide an unequivocal means of determining *in vitro* specificity. In such cases, detailed metabolite profiling of the tissues expressing the particular enzyme in question becomes an essential tool for clarification of function.

Experimental procedures

Mining of EST databases

Medicago truncatula EST databases at TIGR (ww.tigr.org) and in-house at the Noble Foundation (http://bioinfo.noble.org) were

mined for all possible glycosyltransferase sequences. The search was carried out using the following keywords (based on the existing protein annotations in the databases): glycosyltransferase, glucosyltransferase, UDP-glycosyltransferase, UDP-glucosyltransferase, among others. We also utilized PFAM and BLAST searches using the UDP-binding sequence of known plant family 1 UGTs from *A. thaliana*. GTs annotated as involved in cell wall biosynthesis were discarded. UGTs were included in subsequent analyses only if they contained the carboxy terminal PSPG box consensus sequence (Vogt and Jones, 2000).

Plant material

Medicago truncatula Gaerth cv. Jemalong (line A17) plants were grown under greenhouse conditions in 11 cm diameter pots (nine plants per pot) in Metro-mix 250 or 350 (Scott, Marysville, OH, USA), with an 18-h light/25°C and 6-h dark/22°C photoperiod. Cell suspension cultures were initiated from roots, maintained in a modified Schenk and Hildebrandt medium (Dixon *et al.*, 1981), and subcultured every 10–14 days. Six days after subculture, dark-grown suspension cultures (40 ml batches) were treated with MeJA (500 μM, final concentration), harvested at various time points after elicitation and frozen at –80°C. Control cells were treated with 0.4 ml of 100% ethanol per 40 ml culture.

RNA analysis

RNA was extracted from *M. truncatula* plants by standard methods (Sambrook *et al.*, 1989). Twenty micrograms of *M. truncatula* RNA from cell suspension cultures or plant tissues was separated by electrophoresis in a 1% agarose gel containing 0.66 M formaldehyde and then blotted onto a Hybond-N$^+$ membrane (Amersham, St Louis, MO, USA). The entire cDNA fragments of GTs or β-AS were radiolabeled with [^{32}P]-dCTP using a Ready-To-Go DNA Labeling Beads (-dCTP) kit (Amersham Radiolabeled Chemicals) and used as probes. RNA gel blotting and hybridization were carried out as described (Church and Gilbert, 1984). For reverse transcriptase-polymerase chain reaction (RT-PCR) analysis, 3 μg of total RNA was used in a first-strand synthesis using Ready-To-Go RT-PCR Beads (Amersham Biosciences) in a 50-μl reaction with oligo-dT primers according to the manufacturer's protocol. Three microliters of the first-strand reaction was then PCR-amplified using Takara *Ex Taq* (Fisher Scientific Company, Palatine, IL, USA) according to the manufacturer's protocol for 25 cycles and the PCR products were then analyzed on an agarose gel.

Macroarray analysis

cDNA inserts for all 164 available GT clones in the Noble Foundation EST library collection (in pBluescript) were amplified by PCR using standard M13 forward and reverse primers. Approximately 100 ng of each PCR product was spotted in quadruplicate onto a Hybond-N$^+$ membrane (Amersham). Two membranes were hybridized three times with complex probes prepared from independent total RNA extracts. ^{32}P-Labeled first-strand cDNA probes were made from total RNA from unelicited and 24 h MeJA-elicited cell cultures. Reaction mixtures contained 7 μl of RNA primer solution [3–5 μg of total RNA and 0.5 μg of oligo(dT)$_{12-18}$ primer, annealed by heating to 70°C for 10 min], 4 μl of 5× first strand buffer, 2 μl of 0.1 M dithiothreitol, 1 μl of dNTP mix (2.5 mM dCTP, dGTP, and dTTP, and 0.0625 mM dATP), 5 μl of [α-^{32}P]-dCTP (10 mCi ml^{-1}), and 1 μl (200 units) of SuperScript II reverse transcriptase (Invitrogen Life Technologies,

Carlsbad, CA, USA). After 1 h of labeling at 42°C, 1 μl of 5 mM dCTP was added and the incubation allowed to proceed for a further 30 min. Unincorporated [^{32}P]-dCTP was removed by Sephadex G-50 gel filtration. Membranes were hybridized overnight at 65°C and subsequently washed twice in 5% sodium dodecyl sulfate, 20 mM sodium phosphate, and 1 mM EDTA, pH 8.0, for 10 min each. Radioactive signals were detected by phosphorimaging (Storm 820; Molecular Dynamics/Amersham Biosciences, Piscataway, NJ, USA). Array Vision 6.0 software (IMAGING Research Inc., St Catherines, Ont, Canada) was used for dot identification, dot signal quantification, and local background signal subtraction. *Medicago truncatula* β-AS cDNA (Gene Bank ID CAD23247) was included on the arrays as positive control.

DNA microarray analysis

Microarray analysis using *M. truncatula* 16 000 70-mer oligonucleotide arrays was performed as described previously (Suzuki *et al.*, 2005). Fold-changes in GT transcript levels were determined from analysis of three replicate elicited and three replicate control samples per time point.

Functional identification of recombinant GTs expressed in E. coli

Cloning was performed using standard procedures (Sambrook *et al.*, 1989). *M. truncatula* EST clone NF029H05ST (corresponding to TIGR TC87310, previously TC52812) was PCR amplified from pBluescript II SK+ (Stratagene, La Jolla, CA, USA) with addition of *Bam*HI and *Not*I sites (5'-CGGATCCATGTCTATGAGTGATATAAA-CAAG-3' for the upstream primer and 5'-TGCGGCCGCTCAGTTGC-TTCCTGTAATATCATC-3' for the downstream primer). The PCR product was purified, ligated to pGEMTeasy vector (Promega, Madison, WI, USA), sequenced, excised and re-cloned between the *Bam*HI and *Not*I sites of pET28a(+) (Novagen, Madison, WI, USA). *Escherichia coli* BL21 (DE3) cells harboring the expression construct were grown to an OD$_{600}$ of 0.8, and expression was initiated by addition of isopropyl 1-thio-β-D-galactopyranoside (IPTG) to a final concentration of 0.5 mM, with further incubation with shaking overnight at 20°C.

An identical protocol was used for cloning and expression of EST clone NF049F12LF (corresponding to TIGR TC76351, previously 59098), except that the PCR primers were 5'-CGGATCCAT-GGGAACAGAATCCAAACCATTG-3' for the upstream primer and 5'-TGCGGCCGCCTAGTTTAGCTCCACTGATTT-3' for the downstream primer.

Cell lysates were prepared and His-tagged fusion proteins purified using a His-Tag purification kit (Novagen).

Triterpene GT activity was assayed in a mixture (200 μl) containing 1–3 μg of purified recombinant protein, 14 mM 2-mercaptoethanol, 2.5 mM UDP-Glc or UDP-Gal and 50 mM Tris–HCl, pH 7.0. The acceptor 'substrate' was a mixture of saponin aglycones [medicagenic acid (34%), bayogenin (14.4%), hederagenin (8.9%), soyasapogenol E (0.2%) and three unidentified aglycones (8.2, 7.7 and 1.2%)] purified from *M. truncatula* roots, or a similar preparation from alfalfa (*M. sativa*) roots. Pure hederagenin, soyasapogenol A or soyasapogenol B (1 mM) (ChromaDex, Santa Ana, CA, USA) were also used as acceptors. To test activity of GTs with potential phenolic substrates, the assay mix (200 μl) contained 1–5 μg of recombinant protein, 50 mM Tris–HCl, pH 7.0, 14 mM β-mercaptoethanol, 5 mM UDP-glucose, and 1 mM phenolic substrate mixture containing eight compounds that could be baseline resolved by the HPLC method (see Results). Reactions were carried out at 30°C for 1 h and stopped by

adding 2 volumes of ethyl acetate; after vortexing, the ethyl acetate phase was removed, evaporated to dryness under nitrogen, and the residue resuspended in methanol prior to analysis by HPLC. All enzyme assays were performed at least in duplicate.

Identification of product(s) from assays with triterpene aglycones was performed by HPLC-MS analysis as described previously (Huhman and Sumner, 2002). Products formed by glycosylation of phenolic substrates were analyzed by reverse-phase HPLC on an Agilent HP1100 HPLC equipped with an auto sampler, a quaternary pump, and a diode array detector. Solvent A was 1% aqueous phosphoric acid and solvent B was acetonitrile. The flow rate was 1 ml min^{-1}, isocratic at 5% B for 5 min, followed by a multi-step gradient from 5% B to 10% B in 5 min, 10% B to 17% B in 15 min, 17% B to 23% B in 5 min, 23% B to 50% B in 35 min, and 50% B to 100% B in 4 min. The eluants were monitored at 254, 270, 280, 315, and 524 nm. Identifications were based on chromatographic behavior and UV spectra for aglycones whose concentrations decreased during the enzyme incubation to be replaced with conjugates with similar UV spectra but altered relative mobility. Protein concentrations were determined by the Bradford (1976) method.

Kinetic studies were performed using a reaction mixture containing 50 mM Tris-HCl buffer (pH 7.0), 1 mM 2-mercaptoethanol, 42 µM UDP-[U-^{14}C]-glucose (462.5 Bq, American Radiolabeled Chemicals, Inc.), acceptor substrate (0.01–3 mM) and approximately 3 µg of purified recombinant protein in a final volume of 200 µl. Mixtures were incubated at 30°C for 1 h, extracted twice with 500 µl aliquots of 1-butanol, and radioactivity in a 500-µl aliquot of the organic phase was measured by scintillation counting (Beckman LS6500). V_{max} and K_m values were determined from Lineweaver–Burk plots of initial rate data. K_{cat} values were calculated using the predicted molecular masses of 51 790 and 53 930 g mol^{-1} for GT029H and GT049F, respectively.

GT localization by transient expression of GFP fusion proteins

To construct a C-terminal fusion protein of NF029H (UGT71G1) with sGFP, the UGT71G1 coding region was amplified by PCR with the following primers: forward 5′ACGCGTCGACATGTCTATGAGTGA-TATAAACAAG3′; and reverse 5′ACGCGTCGACGTTGCTTCCTGTA-ATATCATC3′, containing *SalI* sites (underlined). PCR products were recovered, gel-purified, digested with *SalI*, and the UGT71G1 open reading frame inserted into the multiple cloning site of the vector sGFP (S65T) in which the chimeric UGT71G1-sGFP gene was expressed under control of the cauliflower mosaic virus 35S promoter.

Plasmid DNA (approximately 5 µg) harboring UGT71G1-sGFP was mixed with 1.0 µm gold particles for biolistic bombardment. Young *M. truncatula* leaves were excised and placed on moist filter paper in petri dishes. Particle bombardment (Bio-Rad 1000/He particle delivery system; Bio-Rad, Hercules, CA, USA), cellular imaging using a Bio-Rad 1024ES confocal imaging system attached to a Zeiss Axioskop microscope (Carl Zeiss, Thornwood, NY, USA), and collection and processing of serial optical images were performed as described previously (Liu and Dixon, 2001).

A similar genomics approach has recently been used to identify diferpene glycosyltransferases from *Stevia rebandiana* (Richman et al. (2005).

Acknowledgements

Supplementary Material

The following material is available from http://www.blackwellpublishing.com/products/journals/suppmat/TPJ/TPJ2344/TPJ2344sm.htm

Figure S1. Functional characterization of *Medicago* GT049F by expression in *Escherichia coli*.

Figure S2. Activity of recombinant GT029H against phenolic substrates as analyzed by reverse-phase HPLC.

References

Achnine, L., Blancaflor, E.B., Rasmussen, S. and Dixon, R.A. (2004) Co-localization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell*, **16**, 3098–3109.

Anhalt, S. and Weissenböck, G. (1992) Subcellular localization of luteolin glucuronides and related enzymes in rye mesophyll. *Planta*, **187**, 83–88.

Batz, O., Logemann, E., Reinold, S. and Hahlbrock, K. (1998) Extensive reprogramming of primary and secondary metabolism by fungal elicitor or infection in parsley cells. *Biol. Chem.* **379**, 1127–1135.

Behboudi, S., Morein, B. and VillacresEriksson, M.C. (1999) Quillaja saponin formulations that stimulate proinflammatory cytokines elicit a potent acquired cell-mediated immunity. *Scand. J. Immunol.* **50**, 371–377.

Blume, D.E., Jaworski, J.G. and McClure, J.W. (1979) Uridinediphosphate-glucose: isovitexin 7-O-glucosyltransferase from barley protoplasts: subcellular localization. *Planta*, **146**, 199–202.

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

Chan, R.Y.K., Chen, W.F., Dong, A., Guo, D. and Wong, M.S. (2002) Estrogen-like activity of ginsenoside Rg1 derived from *Panax notoginseng*. *J. Clin. Endocrinol. Metab.* **87**, 3691–3695.

Church, G.H. and Gilbert, W. (1984) Genomic sequencing. *Proc. Natl Acad. Sci. USA*, **81**, 65–71.

Dixon, R.A. (2001) Natural products and disease resistance. *Nature*, **411**, 843–847.

Dixon, R.A., Dey, P.M., Murphy, D.L. and Whitehead, I.M. (1981) Dose responses for *Colletotrichum lindemuthianum* elicitor-mediated enzyme induction in French bean cell suspension cultures. *Planta*, **151**, 272–280.

Dixon, R.A., Harrison, M.J. and Paiva, N.L. (1995) The isoflavonoid phytoalexin pathway: from enzymes to genes to transcription factors. *Physiol. Plant.* **93**, 385–392.

Dixon, R.A., Chen, F., Guo, D. and Parvathi, K. (2001) The biosynthesis of monolignols: a 'metabolic grid', or independent pathways to guaiacyl and syringyl units? *Phytochemistry*, **57**, 1069–1084.

Goossens, A., Häkkinen, S., Laakso, I. *et al.* (2003) A functional genomics approach toward the understanding of secondary metabolism in plant cells. *Proc. Natl Acad. Sci. USA*, **100**, 8595–8600.

Hansen, K.S., Kristensen, C., Tattersall, D.B., Jones, P.R., Olsen, C.E., Bak, S. and Moller, B.L. (2003) The *in vitro* substrate regiospecificity of recombinant UGT85B1, the cyanohydrin glucosyltransferase from *Sorghum bicolor*. *Phytochemistry*, **64**, 143–151.

Haridas, V., Higuchi, M., Jayatilake, G.S., Bailey, D., Mujoo, K., Blake, M.E., Arntzen, C.J. and Gutterman, J.U. (2001) Avicins: triterpenoid saponins from *Acacia victoriae* (Bentham) induce

apoptosis by mitochondrial perturbation. *Proc. Natl Acad. Sci. USA*, **98**, 5821–5826.

Hayashi, H., Huang, P. and Inoue, K. (2003) Up-regulation of soyasaponin biosynthesis by methyl jasmonate in cultured cells of *Glycyrrhiza glabra. Plant Cell Physiol.* **44**, 404–411.

Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T. and Saito, K. (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana. Proc. Natl Acad. Sci. USA*, **101**, 10205–10210.

Huhman, D.V. and Sumner, L.W. (2002) Metabolic profiling of saponin glycosides in *Medicago sativa* and *Medicago truncatula* using HPLC coupled to an electrospray ion-trap mass spectrometer. *Phytochemistry*, **59**, 347–360.

Ibrahim, R.K. (1992) Immunolocalization of flavonoid conjugates and their enzymes. In *Phenolic Metabolism in Plants* (Stafford, H.A. and Ibrahim, R.K., eds). New York: Plenum Press, pp. 25–61.

Jones, M. and Elliott, F.C. (1969) Two rapid assays for saponin in individual alfalfa plants. *Crop Sci.* **9**, 688–691.

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

Jones, P., Messner, B., Nakajima, J.I., Schäffner, A.R. and Saito, K. (2003) UGT73C6 and UGT78D1, glycosyltransferases involved in flavonol glycoside biosynthesis in *Arabidopsis thaliana. J. Biol. Chem.* **278**, 43910–43918.

Jung, J.D., Park, H.-W., Hahn, Y., Hug, C.-G., In, D.S., Chung, H.-J., Liu, J.R. and Choi, D.-W. (2003) Discovery of genes for ginsenoside biosynthesis by analysis of ginseng expressed sequence tags. *Plant Cell Rep.* **22**, 224–230.

Kramer, C.M., Prata, R.T.N., Willits, M.G., De Luca, V., Steffens, J.C. and Graser, G. (2003) Cloning and regiospecificity studies of two flavonoid glucosyltransferases from *Allium cepa. Phytochemistry*, **64**, 1069–1076.

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

Kurosawa, Y., Takahara, H. and Shiraiwa, M. (2002) UDP-glucuronic acid: soyasapogenol glucuronosyltransferase involved in saponin biosynthesis in germinating soybean seeds. *Planta*, **215**, 620–629.

Kushiro, T., Shibuya, M. and Ebizuka, Y. (1998) β-Amyrin synthase. Cloning of oxidosqualene cylase that catalyzes the formation of the most popular triterpene among higher plants. *Eur. J. Biochem.* **256**, 238–244.

Latchinian-Sadek, L. and Ibrahim, R.K. (1991) Partial purification and some properties of a ring B-O-glucosyltransferase from onion bulbs. *Phytochemistry*, **30**, 1767–1771.

Li, Y., Baldauf, S., Lim, E.K. and Bowles, D.J. (2001) Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana. J. Biol. Chem.* **276**, 4338–4343.

Lim, E.K., Doucet, C.J., Li, Y., Elias, L., Worrall, D., Spencer, S.P., Ross, J. and Bowles, D.J. (2002) The activity of Arabidopsis glycosyltransferases toward salicylic acid, 4-hydroxybenzoic acid, and other benzoates. *J. Biol. Chem.* **277**, 586–592.

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

Liu, C.-J. and Dixon, R.A. (2001) Elicitor-induced association of isoflavone O-methyltransferase with endomembranes prevents formation and 7-O-methylation of daidzein during isoflavonoid phytoalexin biosynthesis. *Plant Cell*, **13**, 2643–2658.

Moehs, C.P., Allen, P.V., Friedman, M. and Belknap, W.R. (1997) Cloning and expression of solanidine UDP-glucose glucosyltransferase from potato. *Plant J.* **11**, 227–236.

Nowacka, J. and Oleszek, W. (1994) Determination of alfalfa (*Medicago sativa*) saponins by high-performance liquid chromatography. *J. Agric. Food Chem.* **42**, 727–730.

Oleszek, W., Junkuszew, M. and Stochmal, A. (1999) Determination and toxicity of saponins from *Amaranthus cruentus* seeds. *J. Agric. Food Chem.* **47**, 3685–3687.

Osbourn, A.E. (2003) Molecules of interest. Saponins in cereals. *Phytochemistry*, **62**, 1–4.

Papadopoulou, K., Melton, R.E., Leggett, M., Daniels, M.J. and Osbourn, A.E. (1999) Compromised disease resistance in saponin-deficient plants. *Proc. Natl Acad. Sci. USA*, **96**, 12923–12928.

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

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual.* New York: Cold Spring Harbor Laboratory Press.

Schmid, G., Hammer, D.K., Ritterbusch, A. and Grisebach, H. (1982) Appearance and immunohistochemical localization of UDP-glucose: coniferyl alcohol glucosyltransferase in spruce [*Picea abies* (L.) Karst.] seedlings. *Planta*, **156**, 207–212.

Small, E. (1996) Adaptations to herbivory in alfalfa (*Medicago sativa*). *Can. J. Bot.* **74**, 807–822.

Suzuki, H., Achnine, L., Xu, R., Matsuda, S.P.T. and Dixon, R.A. (2002) A genomics approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula. Plant J.* **32**, 1033–1048.

Suzuki, H., Reddy, M.S.S., Naoumkina, M., Aziz, N., May, G.D., Huhman, D.V., Sumner, L.W., Blount, J.W., Mendes, P. and Dixon, R.A. (2005) Methyl jasmonate and yeast elicitor induce differential genetic and metabolic re-programming in cell suspension cultures of the model legume *Medicago truncatula. Planta*, in press.

Tava, A. and Odoardi, M. (1996) Saponins from *Medicago* spp: chemical characterization and biological activity against insects. In *Saponins Used in Food and Agriculture* (Waller, G.R. and Yamasaki, T., eds). New York: Plenum Press, pp. 97–109.

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

Waller, G.R., Jurzysta, M. and Thorne, R.L.Z. (1993) Alleopathic activity of root saponins from alfalfa (*Medicago sativa* L.) on weeds and wheat. *Bot. Bull. Acad. Sin.* **34**, 1–11.

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

Wein, M., Lavid, N., Lunkenbein, S., Lewinsohn, E., Schwab, W. and Kaldenhoff, R. (2002) Isolation, cloning and expression of a multifunctional O-methyltransferase capable of forming 2,5-dimethyl-4-methoxy-3(2H)-furanone, one of the key aroma compounds in strawberry fruits. *Plant J.* **31**, 755–765.

Xu, Y. (2001) Perspectives on the 21st century development of functional foods: bridging Chinese medicated diet and functional foods. *Int. J. Food Sci. Technol.* **36**, 229–242.

Yazaki, K., Inushima, K., Kataoka, M. and Tabata, M. (1995) Intracellular localization of UDPG:p-hydroxybenzoate glucosyltransferase and its reaction product in *Lithospermum* cell cultures. *Phytochemistry*, **38**, 1127–1130.

Zimmer, D.E., Pedersen, M.W. and McGuire, C.F. (1967) A bioassay for alfalfa saponins using the fungus, *Trichoderma viride* pers.ex fr. *Crop Sci.* **7**, 223–224.