Saponin Biosynthesis in *Saponaria vaccaria*. cDNAs Encoding β-Amyrin Synthase and a Triterpene Carboxylic Acid Glucosyltransferase[1][OA]

Dauenpen Meesapyodsuk, John Balsevich, Darwin W. Reed, and Patrick S. Covello*

Plant Biotechnology Institute, Saskatoon, Saskatchewan, Canada S7N OW9

Saponaria vaccaria (Caryophyllaceae), a soapwort, known in western Canada as cowcockle, contains bioactive oleanane-type saponins similar to those found in soapbark tree (*Quillaja saponaria*; Rosaceae). To improve our understanding of the biosynthesis of these saponins, a combined polymerase chain reaction and expressed sequence tag approach was taken to identify the genes involved. A cDNA encoding a β-amyrin synthase (SvBS) was isolated by reverse transcription-polymerase chain reaction and characterized by expression in yeast (*Saccharomyces cerevisiae*). The SvBS gene is predominantly expressed in leaves. A *S. vaccaria* developing seed expressed sequence tag collection was developed and used for the isolation of a full-length cDNA bearing sequence similarity to ester-forming glycosyltransferases. The gene product of the cDNA, classified as UGT74M1, was expressed in *Escherichia coli*, purified, and identified as a triterpene carboxylic acid glucosyltransferase. UGT74M1 is expressed in roots and leaves and appears to be involved in monodesmoside biosynthesis in *S. vaccaria*.

Saponaria vaccaria (synonyms *Vaccaria segetalis*, *Vaccaria hispanica*; Caryophyllaceae) is an annual herb widely distributed in Asia, Europe, and other parts of the world. The seeds of this plant are known in traditional Chinese medicine as Wang-Bu-Liu-Xing, which is prescribed for the treatment of amenorrhea, breast infections, and the stimulation of lactation. Phytochemical investigations of *S. vaccaria* seeds and other tissues have revealed the presence of triterpene saponins based on oleanane-type aglycones (Jia et al., 2002; Sang et al., 2003; Balsevich et al., 2006) similar to those found in the soapbark tree (*Quillaja saponaria*, Rosaceae; Guo et al., 1998).

Triterpenoid saponins are a large class of natural products present in higher plants (Sparg et al., 2004). They exhibit a wide variety of both structural diversity and biological activity. Generally speaking, the biological role of saponins in plants is not very clear, but they are implicated as antimicrobials and antifeedants (Hostettmann and Marston, 1995). In addition, some of these molecules have potentially useful pharmacological activities, including immunogenic, anticholesterolemic, and anticancer activities. Indeed, saponins similar in structure to those found in *S. vaccaria* have found use as adjuvants in vaccines. In spite of the numerous studies concerning the occurrence, chemical structure, and biological activities of saponins, the enzymes and genes involved in the biosynthesis of these complex molecules are largely uncharacterized (Haralampidis et al., 2002).

The saponins of the Caryophyllaceae family, such as those of *S. vaccaria*, are almost completely based on β-amyrin (see Fig. 1). The most common aglycones found in this family are quillaic acid, gypsogenic acid, and gypsogenin, which have hydroxy and carboxylate groups at C-3 and C-28, respectively (Jia et al., 2002). In *S. vaccaria*, the saponins can be divided into two groups, the monodesmosides that contain one ester-linked oligosaccharide, typically at C-28 of gypsogenic acid (e.g. vaccaroside B; Fig. 1; Koike et al., 1998) and the bisdesmosides that contain acetal- and ester-linked oligosaccharides, typically at C-3 and C-28, respectively, of quillaic acid (e.g. vaccaroside E; Fig. 1; Jia et al., 2002).

While relatively little is known about the later stages of saponin biosynthesis in *S. vaccaria*, the likely route to both mono- and bisdesmosides is represented in Figure 1. This is based on biochemical and molecular genetic work from other species (Abe et al., 1993; Hostettmann and Marston, 1995; Mackenzie et al., 1997; Vogt and Jones, 2000; Haralampidis et al., 2002). It is quite possible that some of the steps in the pathway do not occur in the order shown. Furthermore, when all of the saponins found in *S. vaccaria* are considered, a relatively complex metabolic network must be involved. As indicated in Figure 1, the first committed step in the pathway toward saponins is the cyclization of 2,3-oxidosqualene by β-amyrin synthase (BAS), one member of a family of oxidosqualene

[1] This work was supported by the National Research Council of Canada's Crops for Enhanced Human Health Program and Genomics and Health Initiative; NRCC Publication Number 48415.

* Corresponding author; e-mail patrick.covello@nrc-cnrc.gc.ca; fax 306–975–4839.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Patrick S. Covello (patrick.covello@nrc-cnrc.gc.ca).

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.106.088484

cyclases (OSCs; Abe et al., 1993; Haralampidis et al., 2002). Judging from the structure of the saponins that accumulate in *S. vaccaria*, the next steps in the pathway presumably include: (1) oxidation of β-amyrin at positions 16, 23, and/or 28; (2) glycosylation at position 28 and, for the major bisdesmosides, position 3; and (3) the acylation of sugars with acetyl and 2-hydroxy-2-methylglutaryl moieties. Apart from the obvious structural prerequisites (C-28 must be oxidized prior to glycosylation), little is known about the order of the reactions involved. As an example, studies in *Calendula officinalis* (Asteraceae) suggest that the sapogenin oleanolic acid (having a carboxyl at C-28) is formed by stepwise oxidation prior to glycosylation at C3 (Wilkomirski and Kasprzyk, 1979). It is possible that this is a general feature of saponin biosynthesis.

The enzymes involved in oxidation of β-amyrin may include cytochrome P450s and other hydroxylases,

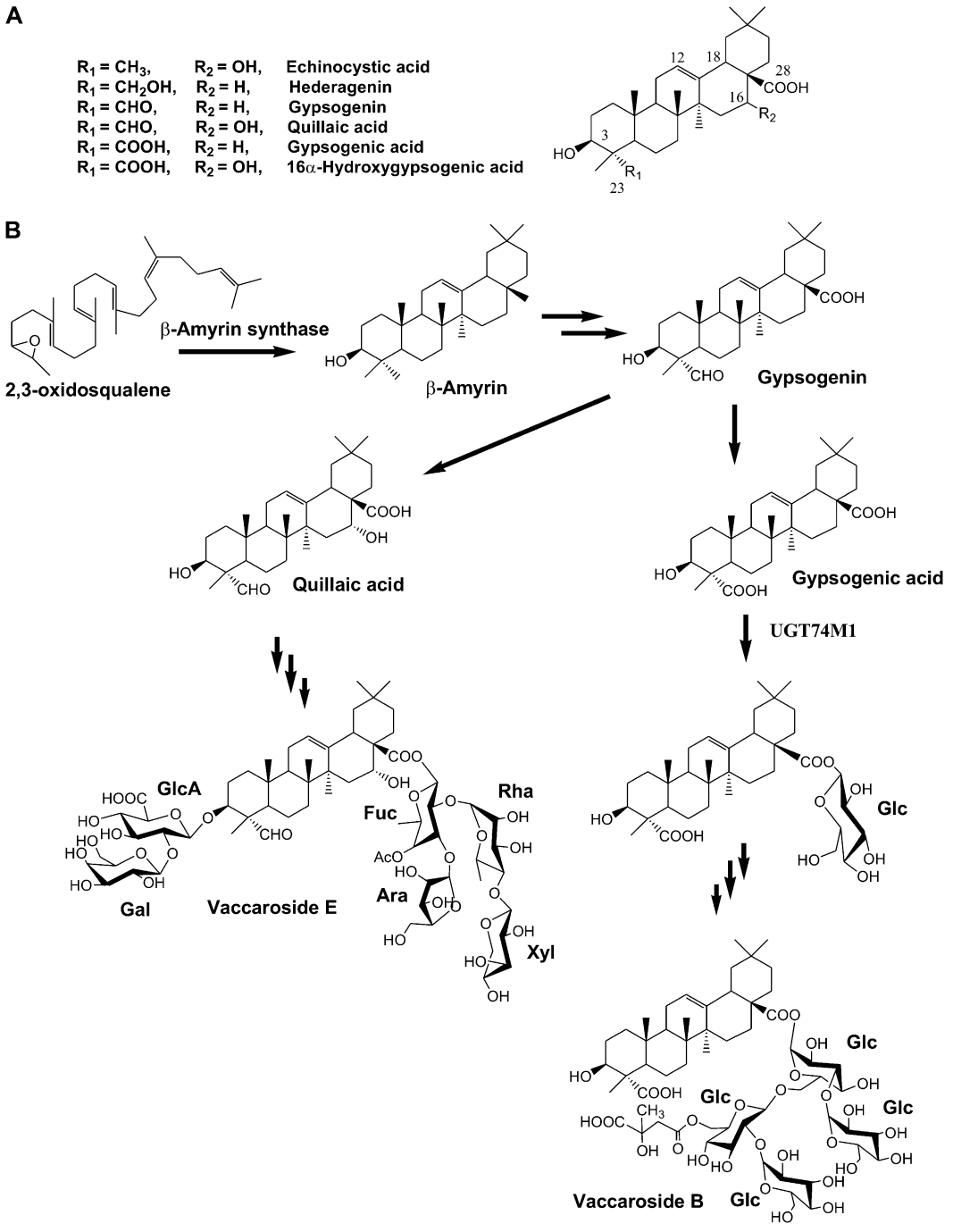

Figure 1. Structure and biosynthesis of *S. vaccaria* saponins. A, Sapogenins of *S. vaccaria* and related compounds. B, Proposed biosynthesis in *S. vaccaria* of examples of a monodesmoside (vaccaroside B) and a bisdesmoside (vaccaroside E).

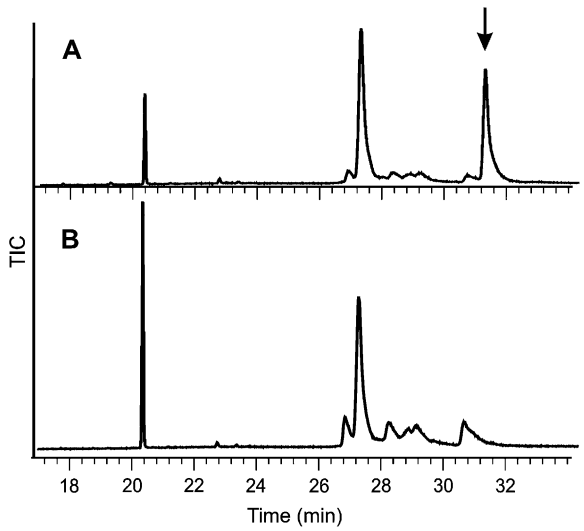

Figure 2. GC-MS analysis of the underivatized product of SvBS expressed in *S. cerevisiae*. Total ion chromatograms are shown for extracts of the *S. cerevisiae* strain MKP-0/pDM067, expressing SvBS (A) and the control strain MKP-0/pSCW231 (B). See "Materials and Methods" for details. Chromatograms represent equal volumes of *S. cerevisiae* cultures. Arrow indicates peak identified as β-amyrin.

and alcohol and aldehyde dehydrogenases. Monodesmosides are presumably formed by stepwise transfer of activated monosaccharide donors. For *S. vaccaria*, the first transfer would be expected to be Glc to the carboxyl at C-28. As in a wide variety of glycosyltransferase reactions in nature, this reaction is likely to be catalyzed by an activated monosaccharide such as UDP-Glc, in this case forming an ester linkage (Vogt and Jones, 2000; Li et al., 2001). Additional sugars would then be transferred and in some cases acylated. For example, vaccaroside B has three additional Glc moieties, one of which is esterified with 2-hydroxy-2-methylglutarate (Fig. 1). Similarly, more complex schemes may be postulated for bisdesmosides.

As part of a broader study of the biochemical genetics of saponin biosynthesis in *S. vaccaria*, we report progress in understanding two of the steps shown in Figure 1 involved in monodesmoside formation through the identification and characterization of cDNAs encoding BAS and an ester-forming triterpene glucosyltransferase.

RESULTS

S. vaccaria BAS

Our investigation of the molecular genetics of saponin biosynthesis in *S. vaccaria* included BAS, which catalyzes the first committed step in the pathway. Degenerate oligonucleotide primers based on known plant BAS genes were used in reverse transcription (RT)-PCR experiments with RNA from germinating seeds as the template. The RACE method was used to obtain a full-length cDNA, as pDM057, corresponding to the gene designated *SvBS*. The *SvBS* open reading frame (ORF) consists of 2,283 nucleotides encoding a 760-amino acid protein of 87.5 kD. The *SvBS* amino acid sequence revealed 81%, 80%, and 72% identity with the BAS of *Glycyrrhiza glabra* (GenBank accession no. AB037203), *Medicago truncatula* (GenBank accession no. AJ430607), and Arabidopsis (*Arabidopsis thaliana*; GenBank accession no. NM106544), respectively. SvBS possesses the amino acid motif DCTAE, thought to form part of the active site of OSC (Abe et al., 1993; Abe and Prestwich, 1995) and the four QW motifs characteristic of the OSC superfamily (Poralla et al., 1994). In addition, SvBS amino acid sequence contains the Trp residue in the MWCYCR motif that plays an important role in the formation of β-amyrin in *Panax ginseng* (Kushiro et al., 2000).

The identity of the enzyme encoded by *SvBS* was confirmed by expression in yeast (*Saccharomyces cerevisiae*). Figure 2 shows the results of gas chromatography-mass spectrometry (GC-MS) analysis of extracts of the yeast strain MKP-0/pDM067. When compared with the control strain, MKP-0/pSCW231, extracts showed a single additional compound whose retention time and mass spectrum (GC-MS[electron impact {EI}] 70 eV, mass-to-charge ratio [m/z; relative intensity]: 218 [100], 203 [52], 69 [25], 95 [25], 55 [24], 135 [18], 189 [17], 426 $[M]^+$ [3], 257 [2], 411 $[M\text{-}CH_3]^+$ [1]) are indistinguishable from authentic β-amyrin (Fig. 2). Thus, the *SvBS* gene product appears to be a BAS that presumably acts on the 2,3-oxidosqualene endogenous to yeast. Expression analysis by RT-PCR indicates that the *SvBS* gene is highly expressed in leaf and to a lesser extent in roots and germinating seeds (Fig. 3). The sequence of the *SvBS* cDNA was deposited in GenBank as accession number DQ915167.

A Triterpene Glucosyltransferase from *S. vaccaria*

The latter stages of saponin biosynthesis involve glycosylation of specific sapogenins. In *S. vaccaria*, the pathway appears to be split into two major routes

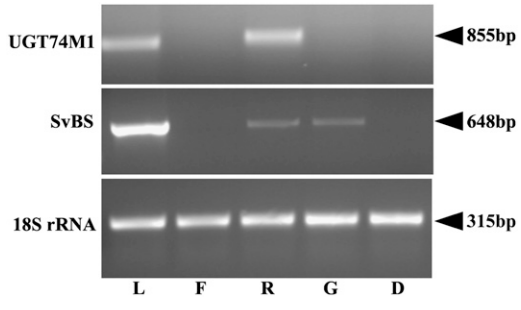

Figure 3. Expression of *SvBS* and *UGT74M1* relative to 18S rRNA in tissues of *S. vaccaria*. Results of RT-PCR are shown for RNA isolated from leaves (L), flowers (F), roots (R), germinating seeds (G), and developing seeds (D). See "Materials and Methods" for details. The sizes in base pairs of the amplified products are shown on the right.

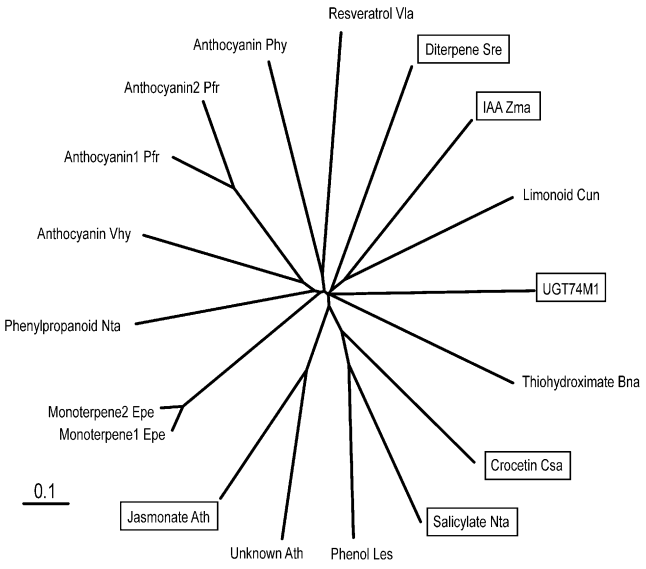

Figure 4. Phylogenetic analysis of UGT74M1 and related sequences. An unrooted distance tree shows the relationship of UGT74M1 to various other glycosyltransferases. The function, species, and GenBank accession numbers of the sequences used are indicated in Table I. Boxes indicate ester-forming UGTs.

directed toward mono- and bisdesmosides (Fig. 1). These are differentiated by both the identity of the aglycone and of the sugar esterified at C-28. The latter is predominantly Glc in monodesmosides and Fuc in bisdesmosides. To investigate this part of the pathway, a search was made for cDNAs encoding glycosyltransferases that could play a role. The rationale behind this expressed sequence tag (EST)-based approach is that most of the ESTs corresponding to genes involved in saponin biosynthesis will have sequences that readily reflect their enzyme class (e.g. glycosyltransferases). In addition, for compounds such as saponins, which are abundant in the tissue of interest, the relevant genes would be expected to be expressed at moderate to high levels and therefore be represented in moderately sized EST collections. Similar approaches have been used to isolate cDNAs encoding other enzymes of secondary metabolism (van de Loo et al., 1995; Cahoon and Kinney, 2004).

Analysis of 7,200 ESTs from a *S. vaccaria* developing seed library indicated that 10 ESTs in four groups showed similarity to plant glycosyltransferases containing the plant secondary product glucosyltransferase domain (Vogt and Jones, 2000). Given its similarity to gene-encoding, ester-forming glucosyltransferases (see below), pSv33B05, a singleton full-length cDNA representing one of the four groups, was investigated as a candidate for involvement in C-28 glycosylation in saponin biosynthesis (the investigation of the additional glycosyltransferase cDNAs, as well as other cDNAs, as candidates involved in saponin biosynthesis is in progress). The *S. vaccaria* gene corresponding to pSv33B05 was classified by Peter Mackenzie (Mackenzie et al., 1997) and given the name *UGT74M1* (and the specific allele *UGT74M1-1*; see below).

DNA sequence analysis of pSv33B05 revealed an ORF corresponding to 478 amino acids and a predicted molecular mass of 53.3 kD. Southern hybridization results indicate that a single copy of *UGT74M1* gene is present in the genome of *S. vaccaria* (data not shown). To identify possible introns in this gene, genomic DNA of *S. vaccaria* was used as a template for PCR with gene-specific primers corresponding to the 5′- and 3′-untranslated regions of the *UGT74M1* gene. A product larger than expected from the cDNA was cloned into the vector pCR2.1-TOPO-TA and sequenced. It was found that this clone contained one intron of 354 bp corresponding to positions 712 to 1,065 in the genomic DNA sequence obtained (deposited in

Table I. *Glycosyltransferase amino acid sequences used in Figure 4*
Ester-forming UGTs are indicated in bold.

Label	Function	Species	GenBank No.
Anthocyanin Vhy	Anthocyanin 5-*O*-glucosyltransferase	*Verbena* × *hybrida*	BAA36423
Anthocyanin Phy	Anthocyanin 5-*O*-glucosyltransferase	*Petunia* × *hybrida*	BAA89009
Anthocyanin1 Pfr	Anthocyanin 5-*O*-glucosyltransferase	*Perilla frutescens*	BAA36422
Anthocyanin2 Pfr	Anthocyanin 5-*O*-glucosyltransferase	*P. frutescens*	BAA36421
Crocetin Csa	**Crocetin glucosyltransferase**	***Crocus sativa***	**AAP94878**
Diterpene Sre	**Steviol glucosyltransferase**	***S. rebaudiana***	**AAR06920**
IAA Zma	**Indole-3-acetate glucosyltransferase**	***Zea mays***	**Q41819**
Jasmonate Ath	**Jasmonic acid glucosyltransferase**	**Arabidopsis**	**ABA39729**
Limonoid Cun	Limonoid glucosyltransferase	*Citrus unshiu*	Q9MB73
Monoterpene1 Epe	Monoterpene glucosyltransferase	*Eucalyptus perriniana*	BAD90935
Monoterpene2 Epe	Monoterpene glucosyltransferase	*E. perriniana*	BAD90934
Phenol Les	UDP-Xyl:phenolic glycosyltransferase	*Lycopersicon esculentum*	CAI62049
Phenylpropanoid Nta	Phenylpropanoid *O*-glucosyltransferase	*N. tabacum*	BAB88935
Resveratrol Vla	Resveratrol/hydroxycinnamic acid *O*-glucosyltransferase	*Vitis labrusca*	ABH03018
Salicylate Nta	**Salicylic acid glucosyltransferase**	***N. tabacum***	**AAF61647**
Thiohydroximate Bna	Thiohydroximate *S*-glucosyltransferase	*B. napus*	AAL09350
Unknown Ath	Unknown	Arabidopsis	NP172059

```
UGT74M1 175 KLPKSFLAYGDNNSHNNNNNNNNNNNNNNMGLHPLVLWLLKD
SA-GT   160 ISIPGLLTIEASDVPSFVSNPESSR---------ILEMLVNQ
UGT74E2 172 DLP-SFLCESSS-------------------YPNILRIVVD
UGT74G1 174 RWETPLIL-----------------QNHEQIQSPWSQMLFGQ
```

Figure 5. Partial amino acid sequence alignment of UGT74M1 and related glycosyltransferases in the region of the polyAsn tract (AA175-216). UGT74M1, from this study; SA-GT, UDP-Glc:salicylic acid glucosyltransferase from *N. tabacum* (GenBank accession no. AAF61647); UGT74E2, indole-3-acetate β-glucosyltransferase from Arabidopsis (GenBank accession no. AAD30627); UGT74G1, UDP-glycosyltransferase 74G1 from *S. rebaudiana* (GenBank accession no. AAR06920).

GenBank as accession no. DQ915168). The position and phase of this intron matches that of introns in Arabidopsis genes corresponding to a subset of plant glucosyltransferases that has been called cluster L of family 1 (Li et al., 2001).

The relationship of UGT74M1 to other plant glycosyltransferases was also assessed through phylogenetic analysis (Fig. 4; Table I) of deduced amino acid sequences. *UGT74M1* was found to lie within a clade that includes members of family 1, cluster L. A number of UDP-glycosyltransferases (UGTs) in this cluster are known to form ester or sulfur linkages. The known enzymes that showed the highest amino acid sequence similarity include *Nicotiana tabacum* salicylic acid glucosyltransferase (GenBank accession no. AAF61647, 39% identity), *Brassica napus* thiohydroximate glucosyltransferase (GenBank accession no. AAL09350, 38% identity), and *Stevia rebaudiana* UGT74G1 (GenBank accession no. AAR06920, 37% identity). This suggests a possible role for UGT74M1 in ester formation and, in the context of known and abundant compounds from *Saponaria* spp., the formation of hexose esters at C-28 of sapogenins, such as gypsogenic acid (see Fig. 1).

An unusual observation regarding the predicted amino acid sequence of UGT74M1 is the presence of 14 contiguous Asn residues. Based on sequence alignments, this polyAsn tract does not appear to share homology with other plant UGTs (see Fig. 5). Indeed, the nucleotide sequence corresponds to the repeated trinucleotide AAT, i.e. a simple sequence repeat. Such sequences are frequently polymorphic in plant populations. To investigate this further, cDNA and genomic *UGT74M1* clones were isolated using PCR. Twelve clones derived from cDNA were sequenced and found to contain nine, 11, 12, 13, and 14 Asn codons with frequencies of 1, 2, 3, 1, and 5, respectively. Four clones from genomic DNA yielded polyAsn tracts of 14 (three clones) and 11 (one clone). Thus, the *UGT74M1* gene appears to be polymorphic within the seed lot used. While the above observations could result from polyploidy, *S. vaccaria* is reported to be diploid (Khoshoo and Bhatia, 1961).

To characterize the activity of UGT74M1, the insert of pSv33B05 was subcloned into the vector pET14b for expression in *Escherichia coli*. Glycosyltransferase activity was determined in cell-free extracts using radiolabeled substrates (UDP-[U-^{14}C]-Glc and UDP-[U-^{14}C]-GlcUA). Preliminary assays showed that the UGT74M1-1 gene product has activity with sapogenin mixture extracted from *S. vaccaria* mature seeds with UDP-[U-^{14}C]-Glc (data not shown). No activity was found when UDP-[U-^{14}C]-GlcUA was used.

To further characterize the properties of this enzyme, recombinant UGT74M1 was purified by immobilized metal affinity chromatography and gel filtration (Fig. 6). The purity was judged to be greater than 80%. The yield of purified UGT74M1 was approximately 1 mg/L culture. Using a variety of triterpene acceptors, including a sapogenin mixture from *S. vaccaria*, β-amyrin, quillaic acid, and oleanolic acid, the purified enzyme was found to be inactive with UDP-GlcUA and GDP-Fuc. Conversely, UDP-Glc was a donor, and the corresponding acceptor specificity of UGT74M1 was determined using various types of saponin aglycones that are present in *S. vaccaria* or available commercially. As shown in Table II and Figure 7, this recombinant enzyme recognized gypsogenic acid, 16α-hydroxygypsogenic acid, quillaic acid, gypsogenin, hederagenin, echinocystic acid, and betulinic acid as acceptor substrates. In contrast, the other oleanane triterpenes β-amyrin, oleanolic acid, and erythrodiol and a variety of other substrates were not converted by UGT74M1 (data not shown; see Table II legend). Gypsogenic acid was used to determine the temperature and pH optima for UGT74M1-1 of 30°C and 7.5,

Table II. *The substrate specificity of UGT74M1*

UGT74M1 assays were performed as described in "Materials and Methods" using the radiochemical assay. The substrates for which no activity was detected are α-amyrin, β-amyrin, asiatic acid, benzoic acid, caffeic acid, cholesterol, cyanidin, diosgenin, erythrodiol, lupeol, oleanolic acid, quercetin, salicylic acid, and spinasterol.

Substrate	Relative Activity	C-23	C-16
	%		
16α-Hydroxygypsogenic acid	177	COOH	OH
Gypsogenic acid	100	COOH	H
Gypsogenin	22	CHO	H
Quillaic acid	19	CHO	OH
Echinocystic acid	9	CH_3	OH
Hederagenin	4.5	CH_2OH	H
Betulinic acid	2.5	CH_3	H

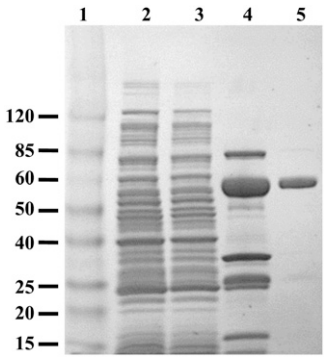

Figure 6. Expression and purification of UGT74M1. SDS-PAGE of extracts of the *E. coli* strain Rosetta2pLysS/pDM064 are shown (see "Materials and Methods" for details). Lane 1, Molecular mass standards (masses in kilo Dalton shown on left); lane 2, crude extract of Rosetta2pLysS/pDM064; lane 3, flowthrough from HisTrap FF Crude column; lane 4, imidazole eluate of HisTrap FF Crude column; lane 5, UGT74M1-containing HPLC fraction.

respectively. Although the enzyme was routinely assayed with 10 mM $MgCl_2$, this was not required for activity (data not shown).

The kinetic constants shown in Table III indicate that 16-hydroxygypsogenic acid was the most efficiently converted to glucoside by the *UGT74M1* gene product. Gypsogenin and quillaic acid had apparent K_m values comparable to 16-hydroxygypsogenic acid but lower k_{cat} values. The presumed major substrate in *S. vaccaria*, gypsogenic acid, had a k_{cat} value similar to 16-hydroxgypsogenic acid but a higher apparent K_m.

In general, glucosylation of the sapogenins can lead to the formation of Glc esters or acetals. The two types of reaction products can be distinguished by alkaline hydrolysis. The products obtained from the enzyme assay using a variety of *S. vaccaria*-derived sapogenin substrates were found to be unstable in the presence of 1 N KOH at 80°C for 2 h (data not shown). This indicates that the product of the enzyme is a Glc ester (see Fig. 1). This was confirmed for the product of gypsogenin glucosylation by NMR (see Table IV). The measured ^{1}H-NMR spectrum of gypsogenin has signals at chemical shifts of 3.30 and 5.48 ppm, which, based on previous NMR studies of gypsogenin glycosides (Delay et al., 1997; Bouget-Bonnet et al., 2002), can be assigned to C-18 and C-12, respectively. In the same region of the spectrum of the UGT74M1-glucosylated product of gypsogenin, signals are also present at 3.20 and 5.48 ppm. In addition, resonances corresponding to a Glc moiety (and C-3) are apparent in the 4.0 to 4.5 ppm range and at 6.36 ppm (1H, d, 8.1 Hz), the latter of which is characteristic of C-1 in Glc esters. Thus, the NMR is consistent with the glucosylation of gypsogenin at the carboxyl group (C-28; see Fig. 1). The UGT74M1 variant derived from pDM066, which lacked the polyAsn tract entirely, was found to exhibit similar glucosyltransferase activity using gypsogenic acid (data not shown).

DISCUSSION

The cloning of cDNAs encoding BAS and UGT74M1 provides some insights into saponin biosynthesis in *S. vaccaria*. The expression of the two genes appears to be tissue specific but not tightly coordinated. For example, some expression of SvBS is observed in germinating seeds for which no *UGT74M1* expression was detected. Based on the observed expression levels for UGT74M1, it is not surprising that it is represented only once in the developing seed EST collection. Thus, the molecular cloning of UGT74M1 reported apparently corresponds to the isolation of a rare cDNA from a rare mRNA.

The characterization of the UGT74M1 product indicates that it is a triterpene carboxylic acid glucosyltransferase. In vitro, the enzyme is capable of glucosylating a variety of oleanane triterpenes as well as having low activity with the lupane triterpenoid, betulinic acid. NMR analysis of the glucosylation product of gypsogenin indicates that it forms a Glc ester at C-28.

It is noteworthy to consider the activity of UGT74M1 in relation to the saponin profile of *S. vaccaria* seeds. The monodesmosides consist primarily of the vaccarosides A to D, having gypsogenic acid as the aglycone and a Glc linked to the carboxyl at C-28. Alternatively, the bisdesmosides (vaccarosides E–H), in addition to GlcUA at C-3, have a Fuc esterified to C-28. Based on our experiments with GDP-Fuc and a variety of aglycones, UGT74M1 does not appear to be involved in making the Fuc ester linkage found in bisdesmosides. It is possible, however, that, in this regard, the correct combination of donor and acceptor was not tested.

Table III. *Kinetic constants for UGT74M1 and various sapogenins substrates*

See "Materials and Methods" for details.

Substrate	K_m^{app}	k_{cat}^{app}	$(k_{cat}/K_m)^{app}$
	μM	s^{-1}	$s^{-1}{}_M{}^{-1}$
Gypsogenic acid[a]	170	1.13	6.5×10^3
16-OH gypsogenic acid	51	1.29	2.5×10^4
Gypsogenin	42	0.125	3.0×10^3
Quillaic acid	37	0.111	3.0×10^3

[a]Constants for gypsogenic acid are considered estimates, because solubility limited the maximum concentration used to 100 μM.

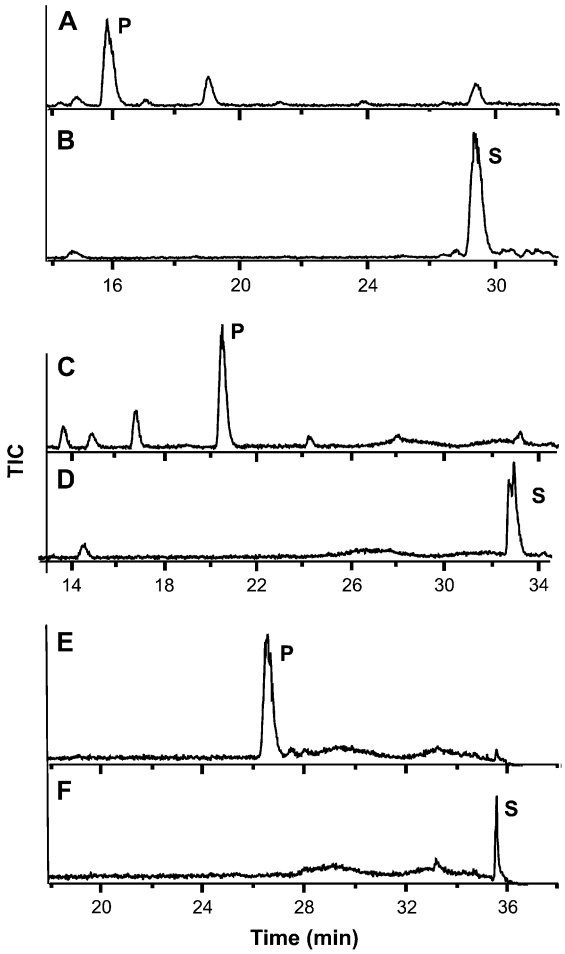

Figure 7. Glucosylation of sapogenins by UGT74M1. Total ion chromatograms from LC-MS are shown for enzyme assays using extracts of the *E. coli* strains Rosetta2pLysS/pDM064 (A, C, and E) and Rosetta2-pLysS/pET14b (B, D, and F) in the presence of 16-hydroxygypsogenic acid (A and B), gypsogenic acid (C and D), and gypsogenin (E and F). Chromatograms represent equal volumes of enzyme assay mixtures. Substrate and major glycosylated product peaks are indicated by S and P, respectively.

Thus, the data presented suggest that UGT74M1 is specific for UDP-Glc and involved in the formation of monodesmosides. It is possible that glucosylation of triterpene carboxylic acids represents a branchpoint in the biosynthesis of mono- and bisdesmosides.

UGT74M1 is a member of cluster L of family 1 GTs. This group includes a number of other ester-forming GTs involved in plant secondary metabolism and hormone metabolism (Li et al., 2001; Lim, 2005). It is notable that the legume *M. truncatula* contains oleanane saponins that have C-28 Glc ester moieties. While UGT71G1 and UGT73K1 have been reported to have activity on hederagenin, it is not clear that they are capable of catalyzing ester formation at C-28 (Achnine et al., 2005). Therefore, it is interesting to speculate that a GT more similar to UGT74M1 and other family 1, cluster L enzymes may be involved in saponin biosynthesis in legumes as well as other members of the Caryophyllaceae.

Clearly, further work is required to elucidate the enzymes and genes involved in other steps in the biosynthesis of saponins in *S. vaccaria*. Our EST collection, in combination with heterologous expression and other experiments, should provide an effective basis with which to uncover the enzymes that catalyze the oxidation, glycosylation, and acyl transfer steps involved. The oxidation of β-amyrin to various sapogenins will be of particular interest and relevance to a number of saponin-producing taxa.

MATERIALS AND METHODS

Chemicals

α-Amyrin, β-amyrin, echinocystic acid, erythrodiol, lupeol, oleanolic acid, and hederagenin were purchased from the Indofine Chemical Company. Asiatic acid, betulinic acid, cholesterol, caffeic acid, diosgenin, quercetin, and salicylic acid were obtained from Sigma-Aldrich. Spinasterol and benzoic acid were obtained from Chromadex and Fisher Scientific, respectively. Cyanidin was obtained from APIN Chemicals. *N*,*O*-bis(trimethylsilyl)-acetamide (BSA) was obtained from Aldrich.

Preparation of Gypsogenin and Mixture of Gypsogenic Acid, 16-Hydroxygypsogenic Acid, and Seco-Gypsogenic Acid from *Saponaria vaccaria* Saponins

Two hundred grams of *Saponaria vaccaria* seed was milled in a coffee grinder. The solids remaining after diethyl ether extraction were air dried and extracted twice with 70% methanol (aq, 700 mL) for 4 h at 50°C. The combined extract was concentrated in vacuo to about 100 mL and applied to an Amberchrom CG-300C (500 g) open column and made up in 10% methanol. The column was eluted with a methanol gradient from 20% to 100% in 20% increments and fractions collected, monitored by thin-layer chromatography, and checked by liquid chromatography (LC)-MS-photodiode array detection (PDA) for composition. Fractions obtained with 60% methanol were enriched in gypsogenic acid saponins such as vaccaroside B (molecular weight [MW] 1,278; see Fig. 1), while fractions obtained with 100% methanol were enriched in the gypsogenin saponin, segetoside H (MW 1,448). The appropriate fractions were combined and evaporated to dryness, affording segetoside H-enriched and vaccaroside B-enriched materials.

Table IV. *Assignment of selected ^{1}H-NMR chemical shifts (given in parts per million) in the 3 to 6.5 ppm range for the gypsogenin glucoside product of UGT74M1*

Atom No.	Gypsogenin 3-*O*-Glucuronide[a]	Gypsogenin	Gypsogenin Glucoside (UGT74M1 Product)
12	5.45	5.48(t)	5.48(t)
18	3.29	3.30(dd)	3.20(dd)
1′	4.87	–	6.36(d)

[a]From Bouget-Bonnet et al. (2002).

Table V. *Oligonucleotide primers used in this study*

Primer Name	Sequence (5′ to 3′)
BS-forward	CGATTTTGTAACACGTACGTACAACACG
BS-reverse	CCCGAGAATTACATGGTAGTAATACAGAAGG
BS-forward1	GAATTCATGTGGAGGTTAAAGATAGCCGAAG
BS-reverse1	GAATTCAGACGAGATTATGGCGATTATTAG
BS-forward2	GCCGTCAACTACCTCGCCGCCATACAAGC
BS-reverse2	CGGGTCTTCAACCCAGCACGCCAACATAC
GT33-F1	CCGCTCGAGATGTCTAATAATGAAAACAATGC
GT33-R1	CCGCTCGAGTTAAGAAGACGAAAGCCACTC
GT33-F4	ATGTCTAATAATGAAAACAATGCAACTCAAGTAATAG
GT33-R4	ACAGAAGAAGCTATCCGACTATCCCTCC
GT33-SEQ3	CAGCAGCGTCCGCCTCGTTTGCTTGTTAC
GT33-SEQ4	GCACGACCATGCCTTGTCCCGAGATTTTC
GT33-SEQ5	ATGTGAATTATTGTCACCATATGC
GT33-SEQ6	GCATATGGTGACAATAATTCACATATGGGCCTTCACCCTTTAG
SQ4	A(A/G)(A/G)TA(I/C)(G/C)(A/T)(C/T)TC(I/C)CCCCA(I/C)CC (I/C)CCI
SQ5	GA(T/C)GGIGGITGGGGI(T/C)TICA
SQ8	(A/T)(C/G)ITT(T/C)(C/T)TICCIATGCA(T/C)CCIGCIAA
SQ9	TC(T/C)TGIGCICCIGCIGG(T/C)TCCCAIGC
SQ10	GATGGACCCTGAGAGACTCTATGAC
SQ11	CAGACCGTTGAATGGCCAGCGAGTAAGAAGAGG
SQ12	GGTGTAATCGGACCAACGAACCTC

Gypsogenin

The segetoside H-enriched material from above (approximately 200 mg) was dissolved in 1.5 M HCl (7 mL) and EtOAc (1 mL) and heated to 90°C for 27 h. The reaction was cooled to ambient temperature and diluted with brine (20 mL). The pH was adjusted to approximately 5 with 1 M NaOH and citric acid, and the mixture was extracted with EtOAc (3 × 25 mL). The combined organic extract was washed with brine (1×), dried (Na_2SO_4), and concentrated in vacuo. The residue was chromatographed on silica gel using diethyl ether as eluant to afford gypsogenin (MW 470) as a white solid, homogeneous by thin-layer chromatography and HPLC. GC-MS of trimethylsilyl (TMS) derivative gypsogenin showed >95% purity. ^{1}H-NMR (pyr.-d_5): δ 9.65 (1H, s), 5.51 (1H, t, 3.4), 4.10 (1H, t, 8.1), 3.34 (1H dd, 13.9, and 4.1), 1.38 (3H, s), 1.30 (3H, s), 1.03 (3H, s), 1.01 (3H, s), 0.98 (3H, s), 0.91 (3H, s).

Sapogenin Diacid Mixture

Vaccaroside B-enriched fractions were combined and evaporated to dryness. The residue (approximately 200 mg) was dissolved in 1 M NaOH and heated at 60°C for 5 h under a nitrogen atmosphere. The reaction mixture was cooled to ambient temperature and the pH adjusted to about 5 with 1 M citric acid. The mixture was diluted with brine (20 mL) and extracted with EtOAc (3 × 25 mL). The combined organic extract was washed with brine (1×), dried (Na_2SO_4), and concentrated in vacuo. The residue was chromatographed on silica gel using EtOAc as eluent to afford a mixture consisting predominantly of gypsogenic acid (MW 486; LC-MS and NMR) as well as minor amounts of seco-gypsogenic acid (MW 486) and 16-hydroxygypsogenic acid (MW 502). ^{1}H-NMR (dimethyl sulfoxide-d6, major component): δ 5.12 (1H, bs), 3.66 (1H, t, 7.7), 2.75 (1H, 5, 10.6), 1.08 (3H, s), 0.91 (3H, s), 0.86 (2 × 3H, s), 0.84 (3H, s), 0.69 (3H, s).

Final Purification of Sapogenins

Final purification of each sapogenin (gypsogenic acid, 16α-hydroxygypsogenic acid, and quillaic acid from the diacid mixture and partially purified gypsogenin) was accomplished by HPLC fractionation using an Agilent 1100 series HPLC with a quaternary pump, diode array detector monitoring at 209 nm, and a fraction collector. A Phenomenex Gemini C18 column (250 × 10 mm, 5 μm particle size) was used with an elution gradient from 22.5% CH_3CN, 0.12% CH_3COOH to 35% CH_3CN, 0.12% CH_3COOH at a flow rate of 3 mL min^{-1} for 30 min. One-minute fractions were collected and the fractions containing detected peaks were analyzed by GC-MS for presence of the sapogenins of interest. The fractions that contained individual sapogenins were combined, and solvent was removed by nitrogen stream and the sapogenins were extracted into ethyl acetate. Trimethylsilyl derivatives of the isolated sapogenins were prepared by addition to a mixture of 50% pyridine, 50% BSA before analysis. The purity of gypsogenic acid (as TMS derivative) GC-MS(EI) 70 eV, m/z (relative intensity): 73 (100), 203 (95), 202 (56), 147 (44), 320 (34), 584 [M-HCOOTMS]$^+$ (32), 129 (25), 585 [M-COOTMS]$^+$ (15), 292 (14), 687 [M-CH_3]$^+$ (11), 495 (9), 467 (8), 381 (3), 377 (3), 612 [M-TMSOH]$^+$ (2), 702 [M]$^+$ (1); 16α-hydroxygypsogenic acid (as TMS derivative) GC-MS(EI) 70 eV, m/z (relative intensity): 73 (100), 147 (8), 201 (7), 129 (5), 275 (12), 318 (12), 790 [M]$^+$ (1), 775 [M-CH_3]$^+$ (1), 700 [M-TMSOH]$^+$ (1), 672 [M-HCOOTMS]$^+$ (1), 610 (1), 583 (1), 493 (1), 393 (1); quillaic acid (as TMS derivative) GC-MS(EI) 70eV, m/z (relative intensity): 73 (100), 187 (25), 143 (24), 129 (17), 275 (12), 585 [M-COOTMS]$^+$ (10), 702 [M]$^+$ (8), 305 (8), 612 [M-TMSOH]$^+$ (1), 393 (3), 495 (2), 687 [M-CH_3]$^+$ (1); and gypsogenin (as TMS derivative) GC-MS(EI) 70eV, m/z (relative intensity): 73 (100), 203 (90), 202 (45), 189 (42), 119 (25), 496 [M-HCOOTMS]$^+$ (17), 320 (17), 307 (6), 614 [M]$^+$ (1), 599 [M-CH_3]$^+$ (1), 407 (1) was 95%, 94%, 82%, and 99%, respectively, by GC-MS.

Plant Materials and Growth Conditions

S. vaccaria cv Pink Beauty seeds were obtained from CN Seeds. Plants were grown under 16 h light (150 μE m^{-2} s^{-1}) at 22°C and 8 h dark at 16°C.

S. vaccaria RNA Isolation and cDNA Library Construction

For relative expression study, total RNA was isolated from field-grown *S. vaccaria*. The RNeasy Plant Mini kit (Qiagen) was used for the total RNA isolation from leaves, flowers, roots, and germinating seeds. For developing seeds, RNA was first isolated by the method of Wang and Vodkin (1994) prior to use of the RNeasy Plant Mini kit. Genomic DNA contamination was eliminated by on-column DNase digestion step with RNase-free DNase set (Qiagen). For cDNA library construction, total RNA was prepared from developing seed of *S. vaccaria* approximately 2 to 4 weeks after flowering. The poly(A$^+$) RNA fraction was isolated (PolyATtract mRNA Isolation system, Promega) and used for cDNA library preparation with a SMART cDNA library construction kit (CLONTECH) according to the manufacturer's instructions using the vector pDNR-LIB. DNA sequencing and EST analysis, including similarity searches using BLAST, was carried out as described previously (Li et al., 2006).

Cloning of Putative BAS cDNA from *S. vaccaria*

Based on the highly conserved amino acid regions of known OSCs, four degenerate oligonucleotide primers were synthesized. The nucleotide sequences of these primers (SQ4, SQ5, SQ8, and SQ9) are shown in Table V. First, PCR with SQ5 and SQ4 primers corresponding to amino acid DGGWGLH and LYSEGWGG, respectively, was carried out for 30 cycles (95°C, 1 min, 50°C, 1 min, 72°C, 1.5 min, and final extension at 72°C, 10 min) using cDNA from 7-to-10-d germinating seeds of *S. vaccaria* as a template. The product of the first PCR was applied to a QIAquick spin column (Qiagen) to remove the primers. Nested PCR was carried out with SQ8 and SQ9 primers corresponding to amino acid SFLPMHPAK and EQAGAPEWA, respectively, with first purified PCR product (5 μL) as a template under the same conditions as the first PCR except the extension time at 72°C for 1 min. The expected size (800–900 bp) fragments were separated by electrophoresis (1.2% agarose gel) and purified using a QIAquick gel extraction kit (Qiagen). This DNA fragment was ligated into a pCR2.1-TOPO vector (Invitrogen). To obtain the entire sequences of *SvBS*, the 5′ and 3′ regions were amplified separately using the Marathon cDNA Amplification kit (BD Biosciences CLONTECH) according to the manufacturer's instructions. The primer SQ10 was used to amplify the 3′ region, and SQ11 and SQ12 were used to amplify the 5′ region. The complete sequences were then amplified using specific primers BS-Forward and BS-Reverse and Vent polymerase (New England Biolabs). The resulting bands were gel purified, cloned into a pCR2.1-TOPO-TA cloning vector (Invitrogen) to give plasmid pDM057, and sequenced. The gene corresponding to pDM057 was designated *SvBS*.

Functional Characterization of SvBS

SvBS was characterized by expression in yeast (*Saccharomyces cerevisiae*). Two oligonucleotide primers (BS-Forward1 and BS-Reverse1), including *Eco*RI sites to facilitate subsequent manipulation, were used to amplify the *SvBS* coding region of pDM057 using Vent polymerase. After treatment with Taq polymerase and dATP, the amplified PCR product was directly ligated into the pCR2.1-TOPO vector (Invitrogen). The plasmid was then digested with *Eco*RI and ligated into pSCW231 yeast expression vector to generate the plasmid pDM067. The DNA sequence of the insert was confirmed to be identical to that of pDM057 and in the sense orientation relative to the ADH1 promoter.

The yeast strain MKP-0 (*MATa can1-100 ade2-1 lys2-1 ura3-52 leu2-3, 112 his3-Δ200 trp1-Δ901*; Covello and Reed, 1996) was separately transformed with pSCW231 and pDM067 by the lithium acetate method (Gietz et al., 1992) and selected on minimal agar plates lacking Trp. The resulting yeast strains were designated MKP-0/pSCW231 and MKP-0/pDM067. For assessment of enzyme activity, recombinant yeast cells were grown until stationary phase in 50 mL at 28°C on minimal medium (synthetic dropout) lacking Trp. MKP-0 yeast containing the empty plasmid vector pSCW231 was used as a negative control.

For analysis of *SvBS* products in yeast, the cells of 50 mL of saturated cultures were collected and saponified with 2 mL 10% KOH/methanol at 80°C for 1 h. After extraction with the same volume of hexane and water, the extract was dried and the residue was dissolved in 100 μL of BSA (Aldrich)/pyridine (1:1). GC-MS analysis was carried out using DB-5MS column (J&W Scientific), as described previously (Li et al., 2006).

DNA Extraction and Southern-Blot Analysis

Genomic DNA was isolated from leaves of *S. vaccaria*, essentially as described previously (Bekesiova et al., 1999). Southern-blot analyses were carried out using standard methods. Ten micrograms of *S. vaccaria* genomic DNA were digested with *Eco*RI, *Eco*RV, or *Hin*dIII (Biolab), resolved on 1% agarose gel, and then transferred to HybondN$^+$ membrane (Amersham Biosciences). This was followed by hybridization at 65°C for 20 h with a 715-bp *Nco*I cDNA probe that had been radiolabeled with α-[^{32}P]dCTP using a Random Primers DNA Labeling kit (Invitrogen).

This fragment containing a partial *SvTGT1* cDNA was obtained from the digestion of pDM060. The filter was washed once in 2× sodium chloride/sodium phosphate/EDTA (SSPE), 0.1% SDS for 10 min then 1× SSPE, 0.1% SDS for 15 and 0.1× SSPE, 0.1% SDS for 10 min at the hybridization temperature. The blot was exposed to Super Rx Fuji Medical x-ray film for 3 d.

RNA Isolation

Total RNA was isolated from field-grown *S. vaccaria*. The RNeasy Plant M ini kit (Qiagen) was used for the total RNA isolation from leaves, flowers, roots, and germinating seeds. For developing seeds, RNA was first isolated by the method of Wang and Vodkin (1994) prior to use of the RNeasy Plant Mini kit. Genomic DNA contamination was eliminated by on-column DNase digestion step with RNase-free DNase set (Qiagen).

Relative Expression of *UGT74M1* and *SvBS* by RT-PCR

To investigate gene expression by RT-PCR, first-strand cDNA was synthesized from 5 μg of total RNA using ThermoScript RT-PCR system in a 20-μL reaction with random primers (Invitrogen) according to the manufacturer's instructions. Two microliters of the first-strand reaction was then used as a template for PCR amplification using Platinum Taq DNA Polymerase (Invitrogen). The specific primers BS-Forward2 and BS-Reverse2 were used for the amplification of *SvBS*, and GT33-SEQ3 and GT33-SEQ4 were used for *UGT74M1*. The 18S PCR primer pair (Universal 18S Internal Standards, Ambion) was used for the amplification as an internal control. The PCR was carried out with 3 min at 95°C, followed by 26 cycles of 30 s at 95°C, 30 s at 65°C, and 40 s at 72°C for SvBS and 18S rRNA and 35 cycles for *UGT74M1*. The PCR products were then analyzed on a 1.5% agarose gel.

Phylogenetic Analysis

Using BLASTP to search public databases maintained at the National Center for Biotechnology Information, amino acid sequences with known function and similarity to UGT74M1 were identified. With software hosted at the European Bioinformatics Institute (Chenna et al., 2003), amino acid sequences encoding glycosyltransferases were aligned using ClustalW using default parameters including the Gonnet scoring matrix, a gap penalty of 10, and a gap extension penalty of 0.2. The resulting alignment was used to generate an unrooted phylogenetic tree using the neighbor joining method. The tree was visualized using TREEVIEW (Page, 1996).

UGT74M1 PolyAsn Variants

To test for genetic variation in the polyAsn tract of UGT74M1, full-length cDNAs were cloned by RT-PCR using total RNA from 7-to-10-d germinating seeds of *S. vaccaria* as a template. The primers GT33-F4 and GT33-R4 (Table V) were used to PCR amplify a DNA fragment of the *UGT74M1* ORF under the following conditions: 3 min at 95°C, followed by 30 cycles of 30 s at 95°C, 60 s at 57°C, and 90 s at 72°C and one cycle of 10 min at 72°C using BD Sprint Advantage PCR kit (CLONTECH). The PCR products were then cloned into pCR2.1-TOPO vector. Twelve individual plasmids were chosen for sequencing. The plasmid that contained the allele with 12 contiguous Asn codons in *UGT74M1* was named pDM065.

To construct a UGT74M1 clone in which the polyAsn tract is deleted, the Gene Spicing by Overlap Extension method (Horton et al., 1990) was used. The primers used for the upper fragment were GT33-F1 and GT33-SEQ5 and for the lower fragment were GT33-R1 and GT33-SEQ6 using pDM060 as the template. The pDM060 was originally from the clone pSv33B05, obtained from a developing seed library of *S. vaccaria*. To obtain the corresponding full-length cDNA, PCR was carried out with the primers GT33-F1 and GT33-R1 and the upper and lower fragments as templates. The resulting PCR product was cloned into pCR2.1-TOPO vector, digested with *Xho*I, and ligated into the *Escherichia coli* expression vector pET14b to generate the plasmid pDM066. To obtain the full-length cDNAs of Asn_{14}- and Asn_{12}-containing *UGT74M1* alleles, PCR was carried out with the primers GT33-F1 and GT33-R1 using pDM060 and pDM065, respectively. The resulting PCR products were cloned into pCR2.1-TOPO and digested with *Xho*I. Both fragments were cloned into pET14b to generate the plasmid pDM064 and pDM065 for Asn_{14}- and Asn_{12}-containing *UGT74M1* alleles, respectively. The DNA sequences of the insert for pDM066, pDM064, and pDM065 were confirmed to be identical to that of the original plasmids and in the sense orientation relative to the T7 promoter.

Expression of and Purification Recombinant UGT74M1

Single colonies of the *E. coli* strain Rosetta2pLysS/pDM064 were used to inoculate 10 mL Luria-Bertani medium containing ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL) at 37°C overnight. The fresh culture was to inoculate 50 volumes of Luria-Bertani medium containing the same antibiotics. Bacteria were grown at 37°C to OD_{600} = 0.5 to 1.0. Induction was achieved by addition of 0.4 mM isopropyl-β-D-thiogalactopyranoside. The

cells were maintained overnight at 30°C, harvested by centrifugation (5,000 rpm, 15 min), and stored at −80°C.

Cell pellets were suspended in buffer A (20 mM sodium phosphate, pH 7.4, 10% [v/v] glycerol, 0.5 M NaCl, 10 mM β-mercaptoethanol) plus 50 mM imidazole, protease inhibitor (Calbiochem), 50 μg/mL RNase A, and 20 μg/mL DNase I, and disrupted using a French press. The cell lysate was centrifuged and the supernatant containing the soluble recombinant enzyme was passed through a 1-mL HisTrap FF Crude column (Amersham Biosciences) prepacked with precharged Ni Sepharose 6 FastFlow. After washing with 20 column volumes, the bound enzyme was eluted with buffer A containing 300 mM imidazole. One-milliliter fractions were collected, and those containing UGT74M1 activity were pooled and concentrated by ultrafiltration (Amicon Ultra-4, 50 K NMWL; Millipore). The protein solution was subsequently applied to Bio-Sil TSK 250 HPLC gel filtration column (300 mm × 7.5 mm; Bio-Rad) at a flow rate of 1 mL min^{-1} using 20 mM HEPES, pH 7.5, and 10% glycerol as eluent. Five hundred-microliter fractions were collected, and those corresponding to OD_{280} peaks were assayed for UGT74M1 activity. The fraction showing the highest activity was frozen in aliquots (50 μL) and stored at −80°C. Protein concentration was determined according to Bradford (1976) using bovine serum albumin as standard.

For electrophoretic analysis, proteins were solubilized in SDS sample buffer (containing 50 mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 0.1 M dithiothreitol, 0.1% bromphenol blue), incubated at 95°C for 5 min, and analyzed by SDS-PAGE on 4% to 15% polyacrylamide gradient gels (Bio-Rad). Protein bands were visualized by staining with GelCode Blue stain reagent (Pierce).

Enzyme Reactions

Unless otherwise stated, glucosyltransferase enzyme reactions were carried out in 100 μL containing 0.6 to 0.8 mg/L of purified recombinant protein, 250 μM UDP-Glc, 1 mM acceptor substrate, 100 mM HEPES, pH 7.5, 15% glycerol, 1 mM dithiothreitol, and 10 mM $MgCl_2$. The reaction mixture was incubated at 30°C for 10 min and stopped by the addition of 1 mL ethyl acetate. For radiochemical assays, the reaction mixture included UDP-[U-^{14}C]Glc (Amersham; final specific activity, 37 kBq/mmol), UDP-[U-^{14}C]GlcUA (Amersham; specific activity, 15.47 GBq/mmol), or GDP-[U-^{14}C]Fuc (Perkin Elmer; specific activity, 7.4 GBq/mmol), and a 500-μL aliquot of ethyl acetate extract was mixed with Aquasol-2 (DuPont) and subjected to scintillation counting with a LKB 1219 Rackbeta liquid scintillation counter.

The pH optimum of UGT74M1 was evaluated using the radiochemical assay and 1 mM gypsogenic acid from pH 5.0 to 10.0 using five buffer systems: 100 mM MES-NaOH, pH 5.0 to 7.0; 100 mM MOPS-NaOH, pH 6.0 to 8.0; 100 mM HEPES-NaOH, pH 6.5 to 8.0; 100 mM Tris-HCl, pH 7.0 to 9.0; and 100 mM 2-(cyclohexylamino)ethanesulfonic acid (CHES)-NaOH, pH 8.5 to 10.0. Similarly, the optimal temperature was evaluated from 20°C to 60°C in 10°C intervals at pH 7.5.

For kinetic studies, the radiochemical assay was used, and substrate concentrations and reaction times were varied for gypsogenic acid (10–100 μM, 10 min), gypsogenin (10–100 μM, 30 min), 16-hydroxygypsogenic acid (10–500 μM, 5 min), and quillaic acid (10–500 μM, 30 min). The kinetic constants were estimated from Lineweaver-Burke plots using the average of triplicate measurements. The k_{cat} values were calculated using the predicted molecular mass of 53,352 g mol^{-1}.

In some cases, unlabeled UDP-Glc was used, and the extracted products were concentrated and subjected to LC-MS (see below). For product analysis by NMR, a 20-mL reaction mixture containing 1 mM gypsogenin was incubated overnight and extracted twice with 50 mL ethyl acetate. After evaporation of the ethyl acetate, the product was purified by HPLC using a Zorbax Extended C-18 column (150 mm × 2.1 mm and 5 μm particle size) maintained at 30°C with an elution gradient from 22.5% CH_3CN, 0.12% CH_3COOH to 35% CH_3CN, 0.12% CH_3COOH over 30 min at a flow rate of 0.2 mL min^{-1}. An eluate fraction corresponding to a peak with a retention time of 23 min was found to contain >95% (LC-MS-PDA) of a compound identified as gypsogenin 28-glucoside. The fraction was evaporated to complete dryness, and after dissolving in pyridine-d_5 (Sigma-Aldrich), the proton NMR spectrum was recorded on a Bruker Avance DRX 500 MHz spectrometer equipped with a CryoProbe.

LC-MS

A 2695 Alliance chromatography system, coupled to a ZQ mass detector and a 2996 photodiode array detector (Waters) was used for LC-MS-PDA analysis. A Waters Sunfire 3.5-μm RP C_{18} 150 × 2.1 mm at 35°C with a flow rate of 0.2 mL/min was used. The binary solvent system consisted of 90:10 (v/v) water:acetonitrile containing 0.12% acetic acid (solvent A) and acetonitrile containing 0.12% acetic acid (solvent B). The gradient program used was: 0 to 3 min, 75:25 A:B; 3 to 28 min, 75:25 to 50:50 A:B; 28 to 31 min, 50:50 to 0:100 A:B; and 31 to 39 min, 0:100 A:B. Mass analysis was performed using negative electrospray ionization under the following conditions: capillary, 2.70 kV; cone ramped, from −10 to −60 V; extractor, −3.50 V; RF lens, −0.7 V.

GC-MS

GC-MS analysis was performed using an Agilent 6890 GC equipped with an autoinjector split 30:1 onto a DB-5MS column (30 M × 0.25 mm i.d., J&W Scientific), which was temperature programmed from 125°C to 300°C at 5°C/min. The column was connected to a mass selective detector (Agilent 5973) operating under standard EI^+ conditions.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers DQ915167 and DQ915168.

ACKNOWLEDGMENTS

We are grateful to Jacek Nowak, Larry Pelcher, Brock Chatson, Greg Bishop, and the PBI Bioinformatics and DNA Technology Units for technical support, Margie Gruber for the kind gift of cyanidin, Peter Mackenzie for classifying UGT74M1, and to Paul Arnison, Irving Ramirez-Erosa, Mark Smith, and Jitao Zou for reviewing the manuscript. This is National Research Council of Canada publication number 48415.

Received August 18, 2006; accepted November 27, 2006; published December 15, 2006.

LITERATURE CITED

Abe I, Prestwich GD (1995) Identification of the active site of vertebrate oxidosqualene cyclase. Lipids **30:** 231–234

Abe I, Rohmer M, Prestwich GD (1993) Enzymatic cyclization of squalene and oxidosqualene to sterols and triterpenes. Chem Rev **93:** 2189–2206

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

Balsevich JJ, Bishop GG, Ramirez-Erosa I (2006) Analysis of bisdesmosidic saponins in *Saponaria vaccaria* L. by HPLC-PAD-MS: identification of new quillaic acid and gypsogenin 3-*O*-trisaccharides. Phytochem Anal **17:** 414–423

Bekesiova I, Nap JP, Mlynarova L (1999) Isolation of high quality DNA and RNA from leaves of the carnivorous plant *Drosera rotundifolia*. Plant Mol Biol Rep **17:** 269–277

Bouget-Bonnet S, Rochd M, Mutzenhardt P, Henry M (2002) Total assignment of ^{1}H and ^{13}C NMR spectra of three triterpene saponins from roots of *Silene vulgaris* (Moench) Garcke. Magn Reson Chem **40:** 618–621

Bradford MM (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

Cahoon EB, Kinney AJ (2004) Dimorphecolic acid is synthesized by the coordinate activities of two divergent delta 12 -oleic acid desaturases. J Biol Chem **279:** 12495–12502

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res **31:** 3497–3500

Covello PS, Reed DW (1996) Functional expression of the extraplastidial *Arabidopsis thaliana* oleate desaturase gene (*FAD2*) in *Saccharomyces cerevisiae*. Plant Physiol **111:** 223–226

Delay C, Gavin JA, Aumelas A, Bonnet PA, Roumestand C (1997) Isolation and structure elucidation of a highly haemolytic saponin from the Merck saponin extract using high-field gradient-enhanced NMR techniques. Carbohydr Res **302:** 67–78

Gietz D, St. Jean A, Woods RA, Schiestl RH (1992) Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res **20:** 1425

Guo S, Kenne L, Lundgren LN, Ronnberg B, Sundquist BG (1998) Triterpenoid saponins from Quillaja saponaria. Phytochemistry **48:** 175–180

Haralampidis K, Trojanowska M, Osbourn AE (2002) Biosynthesis of triterpenoid saponins in plants. Adv Biochem Eng Biotechnol **75:** 31–49

Horton RM, Cai ZL, Ho SN, Pease LR (1990) Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. Biotechniques **8:** 528–535

Hostettmann K, Marston A (1995) Saponins. Cambridge University Press, Cambridge, UK

Jia Z, Koike K, Sahu NP, Nikaido T (2002) Triterpenoid saponins from *Caryophyllaceae* family. *In* A. Rahman, ed, Studies in Natural Products Chemistry, Vol 26. Elsevier Science, Amsterdam, pp 3–61

Khoshoo TN, Bhatia SK (1961) Cytogenetics of the interspecific hybrid, *Vaccaria grandiflora* x *Saponaria vaccaria*. Curr Sci **30:** 327–328

Koike K, Jia Z, Nikaido T (1998) Triterpenoid saponins from Vaccaria segetalis. Phytochemistry **47:** 1343–1349

Kushiro T, Shibuya M, Masuda K, Ebizuka Y (2000) Mutational studies on triterpene synthases: engineering lupeol synthase into β-amyrin synthase. J Am Chem Soc **122:** 6816–6824

Li R, Reed DW, Liu E, Nowak J, Pelcher LE, Page JE, Covello PS (2006) Functional genomic analysis of alkaloid biosynthesis in Hyoscyamus niger reveals a cytochrome P450 involved in littorine rearrangement. Chem Biol **13:** 513–520

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 (2005) Plant glycosyltransferases: their potential as novel biocatalysts. Chemistry (Easton) **11:** 5486–5494

Mackenzie PI, Owens IS, Burchell B, Bock KW, Bairoch A, Belanger A, Fournel-Gigleux S, Green M, Hum DW, Iyanagi T, et al (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. Pharmacogenetics **7:** 255–269

Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput Appl Biosci **12:** 357–358

Poralla K, Hewelt A, Prestwich GD, Abe I, Reipen I, Sprenger G (1994) A specific amino acid repeat in squalene and oxidosqualene cyclases. Trends Biochem Sci **19:** 157–158

Sang S, Xia Z, Lao A, Cao L, Chen Z, Uzawa J, Fujimoto Y (2003) Studies on the constituents of the seeds of *Vaccaria segetalis*. Heterocycles **59:** 811–821

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

van de Loo FJ, Broun P, Turner S, Somerville C (1995) An oleate 12-hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homolog. Proc Natl Acad Sci USA **92:** 6743–6747

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

Wang C-S, Vodkin LO (1994) Extraction of RNA from tissues containing high levels of procyanidins that bind RNA. Plant Mol Biol Rep **12:** 132–145

Wilkomirski B, Kasprzyk Z (1979) Free and ester-bound triterpene alcohols and sterols in cellular subfractions of *Calendula officinalis* flowers. Phytochemistry **18:** 253–255