Saponin Biosynthesis in *Saponaria vaccaria*. cDNAs Encoding β-Amyrin Synthase and a Triterpene Carboxylic Acid Glucosyltransferase\textsuperscript{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. In addition, some of these molecules have potentially useful pharmacological activities, including immunogenic, anticholesterol, 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 gypsojenin, 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...
cyclases (OSC s; 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.
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-CH3]1 [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 2](image-url)

**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.

![Figure 3](image-url)

**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.
 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 1. Glycosyltransferase amino acid sequences used in Figure 4

<table>
<thead>
<tr>
<th>Label</th>
<th>Function</th>
<th>Species</th>
<th>GenBank No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthocyanin Vhy</td>
<td>Anthocyanin 5-O-glucosyltransferase</td>
<td>Verbena × hybrida</td>
<td>BAA36423</td>
</tr>
<tr>
<td>Anthocyanin Phy</td>
<td>Anthocyanin 5-O-glucosyltransferase</td>
<td>Petunia × hybrida</td>
<td>BAA89009</td>
</tr>
<tr>
<td>Anthocyanin1 Pfr</td>
<td>Anthocyanin 5-O-glucosyltransferase</td>
<td>Perilla frutescens</td>
<td>BAA36422</td>
</tr>
<tr>
<td>Anthocyanin2 Pfr</td>
<td>Anthocyanin 5-O-glucosyltransferase</td>
<td>P. frutescens</td>
<td>BAA36421</td>
</tr>
<tr>
<td>Crocetin Csa</td>
<td>Crocetin glucosyltransferase</td>
<td>Crocus sativa</td>
<td>AAP94878</td>
</tr>
<tr>
<td>Diterpene Sre</td>
<td>Steviol glucosyltransferase</td>
<td>S. rebaudiana</td>
<td>AAR06920</td>
</tr>
<tr>
<td>IAA Zma</td>
<td>Indole-3-acetate glucosyltransferase</td>
<td>Zea mays</td>
<td>Q41819</td>
</tr>
<tr>
<td>Jasmonate Ath</td>
<td>Jasmonic acid glucosyltransferase</td>
<td>Arabidopsis</td>
<td>ABA39729</td>
</tr>
<tr>
<td>Limonoid Cun</td>
<td>Limonoid glucosyltransferase</td>
<td>Citrus unshiu</td>
<td>Q9M873</td>
</tr>
<tr>
<td>Monoterpen1 Epe</td>
<td>Monoterpen glycosyltransferase</td>
<td>Eucalyptus periniana</td>
<td>BAD90935</td>
</tr>
<tr>
<td>Monoterpen2 Epe</td>
<td>Monoterpen glycosyltransferase</td>
<td>E. periniana</td>
<td>BAD90934</td>
</tr>
<tr>
<td>Phenol Les</td>
<td>UDP-Xyl:phenolic glucosyltransferase</td>
<td>Lycopersicon esculentum</td>
<td>CA162049</td>
</tr>
<tr>
<td>Phenylpropanoid Nta</td>
<td>Phenylpropanoid O-glucosyltransferase</td>
<td>N. tabacum</td>
<td>BABB98935</td>
</tr>
<tr>
<td>Resveratrol Vfa</td>
<td>Resveratrol/hydroxycinnamic acid O-glucosyltransferase</td>
<td>Vitis labrusca</td>
<td>3ABH39018</td>
</tr>
<tr>
<td>Salicylate Nta</td>
<td>Salicylic acid glucosyltransferase</td>
<td>N. tabacum</td>
<td>AAF61647</td>
</tr>
<tr>
<td>Thiohydroximate Bna</td>
<td>Thiohydroximate S-glucosyltransferase</td>
<td>B. napus</td>
<td>AAL9350</td>
</tr>
<tr>
<td>Unknown Ath</td>
<td>Unknown</td>
<td>Arabidopsis</td>
<td>NPI72059</td>
</tr>
</tbody>
</table>
UGT74M1 is a glucosyltransferase from *N. tabacum* (GenBank accession no. AAF61647); UGT74E2, indole-3-acetate β-glucosyltransferase from *Arabidopsis* (GenBank accession no. AAD30627); UGT74G1, UDP-glucosyltransferase 74G1 from *S. rebaudiana* (GenBank accession no. AAR06920). 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 glucosyltransferases 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-glucosyltransferases (UGTs) in this clade 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* thiophydroximate 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 polyplody, *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-14C]-Glc and UDP-[U-14C]-GlcUA). Preliminary assays showed that the UGT74M1-1 gene product has activity with sapogenin mixture extracted from *S. vaccaria* mature seeds with UDP-[U-14C]-Glc (data not shown). No activity was found when UDP-[U-14C]-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 gypgenic acid, 16α-hydroxygypgenic acid, quillaic acid, gypsogenin, hederagenin, echinocystic acid, and betulinic acid as acceptor substrates. In contrast, the other oleanane triterpenes β-amyrin, oleanolic acid, and eriochrysidol and a variety of other substrates were not converted by UGT74M1 (data not shown; see Table II legend). Gypgenic 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

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Relative Activity</th>
<th>C-23</th>
<th>C-16</th>
</tr>
</thead>
<tbody>
<tr>
<td>16α-Hydroxygypgenic acid</td>
<td>177</td>
<td>COOH</td>
<td>OH</td>
</tr>
<tr>
<td>Gypgenic acid</td>
<td>100</td>
<td>COOH</td>
<td>H</td>
</tr>
<tr>
<td>Gypsogenin</td>
<td>22</td>
<td>CHO</td>
<td>H</td>
</tr>
<tr>
<td>Quillaic acid</td>
<td>19</td>
<td>CHO</td>
<td>OH</td>
</tr>
<tr>
<td>Echinocystic acid</td>
<td>9</td>
<td>CH3OH</td>
<td>H</td>
</tr>
<tr>
<td>Hederagenin</td>
<td>4.5</td>
<td>CH3OH</td>
<td>H</td>
</tr>
<tr>
<td>Betulinic acid</td>
<td>2.5</td>
<td>CH3</td>
<td>H</td>
</tr>
</tbody>
</table>
respectively. Although the enzyme was routinely assayed with 10 mM MgCl₂, 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-hydroxygypsogenic 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 \( ^1H \)-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

<table>
<thead>
<tr>
<th>Substrate</th>
<th>( K_m^{app} )</th>
<th>( k_{cat}^{app} )</th>
<th>( k_{cat}/K_m^{app} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gypsogenic acid</td>
<td>170</td>
<td>1.13</td>
<td>6.5 ( \times 10^3 )</td>
</tr>
<tr>
<td>16-OH gypsogenic acid</td>
<td>51</td>
<td>1.29</td>
<td>2.5 ( \times 10^4 )</td>
</tr>
<tr>
<td>Gypsogenin</td>
<td>42</td>
<td>0.125</td>
<td>3.0 ( \times 10^4 )</td>
</tr>
<tr>
<td>Quillaic acid</td>
<td>37</td>
<td>0.111</td>
<td>3.0 ( \times 10^4 )</td>
</tr>
</tbody>
</table>

*\(^a\)Constants for gypsogenic acid are considered estimates, because solubility limited the maximum concentration used to 100 \( \mu M \).*
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 oleane 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,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 1H-NMR chemical shifts (given in parts per million) in the 3 to 6.5 ppm range for the gypsogenin glucoside product of UGT74M1**

<table>
<thead>
<tr>
<th>Atom No.</th>
<th>Gypsogenin 3-O-Glucuronide&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Gypsogenin Glucoside (UGT74M1 Product)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>5.45</td>
<td>5.48(t)</td>
</tr>
<tr>
<td>18</td>
<td>3.29</td>
<td>3.30(dd)</td>
</tr>
<tr>
<td>1’</td>
<td>4.87</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.36(d)</td>
</tr>
</tbody>
</table>

<sup>a</sup>From Bouget-Bonnet et al. (2002).
The pH was adjusted to approximately 5 with 1M NaOH and citric acid. The reaction was cooled to ambient temperature and diluted with brine. The residue was chromatographed on silica gel using diethyl ether as eluant to afford gypsogenin (MW 470) as a white solid, homogeneous by GC-MS 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), 129 (95), 202 (95), 256 (56), 147 (44), 320 (34), 584 [M-HCOOTMS] (1), 599 [M-CH3] (15), 292 (14), 687 [M-CH3] (11), 495 (9), 467 (8), 381 (3), 377 (3), 612 [M-TMSOH] (2), 702 [M+1] (1); 16α-hydroxygypsogenic acid (as TMS derivative) GC-MS(EI) 70 eV, m/z (relative intensity): 73 (100), 147 (8), 202 (17), 129 (95), 132 (12), 790 [M+1] (1), 755 [M-CH3] (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), 725 (12), 585 [M-HCOOTMS]+ (10), 702 [M+1] (8), 305 (8), 612 [M-TMSOH]+ (1), 393 (3), 495 (2), 687 [M-CH3]+ (1); and gypsogenin (as TMS derivative) GC-MS(EI) 70eV, m/z (relative intensity): 73 (100), 203 (95), 202 (45), 189 (42), 119 (25), 496 [M-HCOOTMS]+ (17), 320 (17), 307 (6), 614 [M+1] (1), 599 [M-CH3]+ (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 (130 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 RNasey 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 RNasey 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 *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 construction with a SMART cDNA library construction kit (CLONTECH) according to the manufacturer’s instructions using the vector pDRN-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 DFGGWGLH1 and LYSGWWG, 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 SFLPDHPK and EQGAQAPG, 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 sequences were then amplified using specific primers BS-Forward1 and BS-Reverse1, including the cloning sites to facilitate subsequent manipulation, were used to amplify the 5’ 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 EcoRI 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 AHD1 promoter.

Functional Characterization of *SvBS*

*SvBS* was characterized by expression in yeast (*Saccharomyces cerevisiae*). Two oligonucleotide primers (BS-Forward1 and BS-Reverse1), including EcoRI 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 EcoRI 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 AHD1 promoter.

The yeast strain MKP-0 (MATα can1-100 ade2-1 his3-1 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 of 2% glucose medium (synthetic dropout) lacking Trp. 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 (Bekesiowa et al., 1999). Southern-blots analyses were carried out using standard methods. Ten micrograms of *S. vaccaria* genomic DNA were digested with EcoRI, EcoRV, or HindIII (Biolab), resolved on 1% agarose gel, and then transferred to Hybond N+ membrane (Amersham Biosciences). This was followed by hybridization at 65°C for 24 h with a 715-bp NcoI cDNA probe that had been radiolabeled with [α-32P]dATP 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 RNAeasy 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 RNAeasy 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 TREVIEW (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 pDM066.

To construct a *UGT74M1* clone in which the polyAsn tract is deleted, the Gene Splicing 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 Xhol, and ligated into the *Escherichia coli* expression vector pET14b to generate the plasmid pDM066. To obtain the full-length cDNAs of Asn<sub>14</sub>-<sub>12</sub> and Asn<sub>12</sub>-<sub>9</sub> 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 Xhol. Both fragments were cloned into pET14b to generate the plasmid pDM064 and pDM065 for Asn<sub>14</sub>-<sub>12</sub> and Asn<sub>12</sub>-<sub>9</sub> containing *UGT74M1* alleles, respectively. The DNA sequences of the insert for pDM064, pDM065, 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 Rosetta2(pLysS) pLysS/pDM064 were used to inoculate 10 mLuria-Bertani medium containing ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL) at 37°C overnight. The fresh culture was inoculated to 50 volumes of Luria-Bertani medium containing the same antibiotics. Bacteria were grown at 37°C to OD<sub>600</sub> = 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% 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) pre-packed 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 corresponding to OD280 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.

Enzyme Reactions

Unless otherwise stated, glucosyltransferase enzyme reactions were carried out in 100 μL containing 0.6 to 0.8 mg/mL of purified recombinant protein, 250 μM UDP-Glc, 1 mM acceptor substrate, 100 mM HEPES, pH 7.5, 15% glycerol, 1 mM diethiothreitol, and 10 mM MgCl2. 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-14C]Glc (Amersham; final specific activity, 37 kBq/μmol), UDP-[U-14C]GlcUA (Amersham; specific activity, 15.47 GBq/μmol), or GDP-[U-14C]Fuc (Perkin Elmer; specific activity, 7.4 GBq/μmol), and a 300-μL aliquot of ethyl acetate extract was mixed with Auquasol-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 3-oxoxygycogenic acid from pH 5.0 to 7.0 with 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 1-(cyclohexylamino)ethanesulfonic acid (CHES)-NaOH, pH 8.5 to 10.0. The optimal temperature was evaluated at 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 3-oxoxygycogenic acid (10-100 μM, 10 min), 3-hydroxy-3-oxoxygycogenic acid (100-500 μM, 5 min), and quillaic acid (100-500 μM, 30 min). The kinetic constants were estimated from Lineweaver-Burke plots using the average of triplicate measurements. The kcat 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 3-oxoxygycogenic acid 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% CH3CN, 0.12% CH3COOH to 35% CH3CN, 0.12% CH3COOH 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-FD) of a compound identified as gypsogenin 28-glucoside. The fraction was evaporated to complete dryness, and after dissolving in pyridine-d5 (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 C18, 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 Amison, 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


Downloaded from on November 11, 2017 - Published by www.plantphysiol.org
Copyright © 2007 American Society of Plant Biologists. All rights reserved.


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


