UDP-glycosyltransferases from the UGT73C subfamily in Barbarea vulgaris catalyse Sapogenin 3-O-glucosylation in Saponin-mediated Insect resistance

Augustin, Jörg Manfred; Drok, Sylvia; Shinoda, Tetsuro; Sanmiya, Kazutsuka; Nielsen, Jens Kvist; Khakimov, Bekzod; Olsen, Carl Erik; Hansen, Esben Halkjær; Poulsen, Vera Kuzina; Ekstrøm, Claus Thorn; Hauser, Thure Pavlo; Bak, Søren

Published in:
Plant Physiology

DOI:
10.1104/pp.112.202747

Publication date:
2012

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
UDP-Glycosyltransferases from the UGT73C Subfamily in *Barbarea vulgaris* Catalyze Sapogenin 3-O-Glucosylation in Saponin-Mediated Insect Resistance\[^{1,}[W][OA]\]

Jörg M. Augustin\(^2\), Sylvia Drok, Tetsuro Shinoda\(^3\), Kazutsuka Sanmiya\(^4\), Jens Kvist Nielsen, Bekzod Khakimov, Carl Erik Olsen\(^5\), Esben Halkjær Hansen, Vera Kuzina\(^5\), Claus Thorn Ekstrøm\(^6\), Thure Hauser\(^5\), and Søren Bak\(^5\)*

Department of Plant Biology and Biotechnology (J.M.A., S.D., B.K., V.K., S.B.), Department of Basic Science and Environment (J.K.N., C.E.O., C.T.E.), Department of Food Science (B.K.), and Department of Agriculture and Ecology (J.K.N., T.H.), University of Copenhagen, 1871 Frederiksberg, Denmark; National Institute of Vegetable and Tea Science, National Agriculture and Food Research Organization, 514–2392 Tsu, Mie, Japan (T.S., K.S.); and Evolva A/S, 2100 Copenhagen, Denmark (E.H.H.)

Triterpenoid saponins are bioactive metabolites that have evolved recurrently in plants, presumably for defense. Their biosynthesis is poorly understood, as is the relationship between bioactivity and structure. *Barbarea vulgaris* is the only crucifer known to produce saponins. Hederagenin and oleanolic acid cellobioside make some *B. vulgaris* plants resistant to important insect pests, while other, susceptible plants produce different saponins. Resistance could be caused by glucosylation of the sapogenins. We identified four family 1 glycosyltransferases (UGTs) that catalyze 3-O-glucosylation of the sapogenins oleanolic acid and hederagenin. Among these, UGT73C10 and UGT73C11 show highest activity, substrate specificity and regiospecificity, and are under positive selection, while UGT73C12 and UGT73C13 show lower substrate specificity and regiospecificity and are under purifying selection. The expression of UGT73C10 and UGT73C11 in different *B. vulgaris* organs correlates with saponin abundance. Monoglucosylated hederagenin and oleanolic acid were produced in vitro and tested for effects on *P. nemorum*. 3-O-\(\beta\)-D-Glc hederagenin strongly deterred feeding, while 3-O-\(\beta\)-D-Glc oleanolic acid only had a minor effect, showing that hydroxylation of C23 is important for resistance to this herbivore. The closest homolog in *Arabidopsis thaliana*, UGT79C5, only showed weak activity toward sapogenins. This indicates that UGT73C10 and UGT73C11 have neofunctionalized to specifically glucosylate sapogenins at the C3 position and demonstrates that C3 monoglucosylation activates resistance. As the UGTs from both the resistant and susceptible types of *B. vulgaris* glucosylate sapogenins and are not located in the known quantitative trait loci for resistance, the difference between the susceptible and resistant plant types is determined at an earlier stage in saponin biosynthesis.

---

1 This work was supported by the Danish Council for Independent Research, Technology, and Production Sciences (grant nos. 09-065899/FTP and 274–06–0370), by the Villum Kann Rasmussen Foundation to Pro-Active Plants, and by a PhD stipend from the Faculty of Life Sciences, University of Copenhagen (to J.M.A.).

2 Present address: Donald Danforth Plant Science Center, St. Louis, MO 63132.

3 Present address: Division of Insect Sciences, National Institute of Agrobiological Sciences, Tsukuba, 305-8634 Ibaraki, Japan.

4 Present address: Department of Bioreourses Engineering, Okinawa National College of Technology, Nago, 905-2192 Okinawa, Japan.

5 Present address: Department of Plant and Environmental Sciences, University of Copenhagen, 1871 Frederiksberg, Denmark.

6 Present address: Department of Biostatistics, University of Southern Denmark, 5000 Odense C, Denmark.

* Corresponding author; e-mail bak@life.ku.dk.

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: Søren Bak (bak@life.ku.dk).

\[^{1[W]}\] The online version of this article contains Web-only data.

\[^{[OA]}\] Open Access articles can be viewed online without a subscription. www.plantphysiol.org/cgi/doi/10.1104/pp.112.202747
cyclases (OSCs) forming the core structures, (2) cytochromes P450 adding the majority of functional groups, and (3) family 1 glycosyltransferases (UGTs) adding sugars. This allows for a vast structural complexity, some of which probably evolved by sequential gene duplication followed by functional diversification (Osborn, 2010). A major challenge is thus to understand the processes of saponin biosynthesis, which structural variants of saponins play a role in defense against biotic antagonists, and how saponin biosynthesis evolved in different plant taxa. This knowledge is also of interest for biotechnological production and the use of saponins as protection agents against agricultural pests as well as for pharmacological and industrial uses as bactericides (De Leo et al., 2006), anticancerogens (Musende et al., 2009), and adjuvants (Sun et al., 2009).

*Barbarea vulgaris* (winter cress) is a wild crucifer from the Cardamineae tribe of the Brassicaceae family. It is the only species in this economically important family known to produce saponins. *B. vulgaris* has further diverged into two separate evolutionary lineages (types; Hauser et al., 2012; Toneatto et al., 2012) that produce different saponins, glucosinolates, and flavonoids (Agerbirk et al., 2003b; Dalby-Brown et al., 2011; Kuzina et al., 2011). Saponins of the one plant type make plants resistant to the yellow-striped flea beetle (*Phyllotreta nemorum*), diamondback moth (*Plutella xylostella*), and other important crucifer specialist herbivores (Renwick, 2002); therefore, it has been suggested to utilize such plants as a trap crop to diminish insect damage (Badenes-Perez et al., 2005). The other plant type is not resistant to these herbivores. *B. vulgaris*, therefore, is ideal as a model species to study saponin biosynthesis, insect resistance, and its evolution, as we can contrast genes, enzymes, and their products between closely related but divergent plant types.

Insect resistance of the one plant type, called G because it has glabrous leaves, correlates with the content of especially hederagenin cellobioside, oleanolic acid cellobioside, 4-epi-hederagenin cellobioside, and gypsogenin cellobioside (Shinoda et al., 2002; Agerbirk et al., 2003a; Kuzina et al., 2009; Fig. 1). These saponins are absent in the susceptible plant type, called P because it has pubescent leaves, which contains saponins of unknown structures and function (Kuzina et al., 2011). The saponins (agycones) of the resistance-causing saponins hederagenin and oleanolic acid cellobioside do not deter feeding by *P. nemorum*, which highlights the importance of glycosylation of saponins for resistance (Nielsen et al., 2010). Therefore, the presence or absence of sapogenin glycosyltransferases could be a determining factor for the difference in resistance between the insect resistant G-type and the susceptible P-type of *B. vulgaris*.

Some *P. nemorum* genotypes are resistant to the saponin defense of *B. vulgaris* (Nielsen, 1997b, 1999). Resistance is coded by dominant R genes (Nielsen et al., 2010; Nielsen, 2012): larvae and adults of resistant genotypes (RR or Rr) are able to feed on G-type foliage and utilize *B. vulgaris* as host plant (de Jong et al., 2009), whereas larvae of the susceptible genotype (rr) die and adult beetles stop feeding on G-type foliage. Larvae and adults of all known *P. nemorum* genotypes can feed on P-type *B. vulgaris* (Fig. 2).

In this study, we asked which enzymes are involved in glucosylation of sapogenins in *B. vulgaris*, whether saponins with a single C3 glucosyl group are biologically active, and whether the difference between the insect resistant and susceptible types of *B. vulgaris* is caused by different glucosyltransferases. We report the identification of two UDP-glycosyltransferases, UGT73C10 and UGT73C11, which have high catalytic activity and substrate specificity and regiospecificity for catalyzing 3-O-glycosylation of the sapogenins oleanolic acid and hederagenin. The products, 3-O-β-D-glucopyranosyl hederagenin and 3-O-β-D-glucopyranosyl oleanolic acid, are predicted precursors of hederagenin and oleanolic acid cellobioside, respectively. The expression patterns of UGT73C10 and UGT73C11 in different organs of *B. vulgaris* correlate with saponin abundance, and monoglucosylated sapogenins, especially 3-O-β-D-glucopyranosyl hederagenin, deter feeding by *P. nemorum*. Our results thus show that glycosylation with even a single glucosyl group activates the resistance function of these sapogenins. However, since the UGTs are present and active in both the insect-resistant and -susceptible types of *B. vulgaris*, we cannot explain the difference in resistance by different glycosylation abilities. Instead, the difference between the susceptible and resistant types must be determined at an earlier stage in saponin biosynthesis.

RESULTS

Identification of a Sapogenin UDP-Glycosyltransferase by Activity-Based Screening of a cDNA Expression Library

To identify enzymes that glycosylate sapogenins (aglycones of saponins) from *B. vulgaris*, a complementary
DNA (cDNA) expression library was generated from *B. vulgaris* var *variegata*, a commercial *B. vulgaris* variety with a saponin profile similar to the insect-resistant G-type. The library was screened by activity assays using UDP-Glc and oleanolic acid as donor and acceptor substrate, respectively. A single cDNA clone was identified, of which the encoded enzyme glucosylated oleanolic acid, as evidenced by comigration with authentic 3-O-Glc oleanolic acid on thin-layer chromatography (TLC) analysis. The clone was designated *BvUGT1* and found to contain a 1,566-bp cDNA with an open reading frame (ORF) of 495 amino acids. BLAST analyses identified *Arabidopsis thaliana* UGT73C5 as its closest homolog. *BvUGT1* has 88% nucleotide identity to *UGT73C5*, and the encoded amino acid sequence, *BvUGT1*, is 83% identical to *UGT73C5*. In addition to oleanolic acid, *BvUGT1* also glucosylated hederagenin and echinocystic acid.

**Identification of *BvUGT1* Homologs in G- and P-Type *B. vulgaris***

Putative *BvUGT1* homologs in the resistant G-type and susceptible P-type were searched by mining a 454 transcriptome data set from the G-type (Kuzina et al., 2011) and the P-type. Based on the identified singlets and contigs, two different full-length ORFs from G-type plants and three from P-type plants were isolated by PCR. The genomic sequences were identified by PCR and shown to be intronless, which is also the case for the seven UGT73Cs in the *A. thaliana* genome (Paquette et al., 2003). Thus, putative *BvUGT1* homologs are not only present in both the G- and P-type *B. vulgaris* genomes, but they are also expressed. The three P-type UGTs were named UGT73C9, UGT73C10, and UGT73C12, and the two G-type sequences were named UGT73C11 and UGT73C13 (Fig. 3), by the UGT nomenclature committee (Mackenzie et al., 1997). The ORFs of the five UGTs each span 1,488 bp and encode proteins consisting of 495 amino acids.

Of the five sequences, *UGT73C11* is most identical to *BvUGT1* from *B. vulgaris* var *variegata*, differing in only three nucleotides, which causes a conservative amino acid substitution of Asp-338 to Glu in *UGT73C11*. Based on a reconstruction of the phylogeny of the UGTs (Fig. 3), UGT73C9 and UGT73C10 from the P-type and UGT73C11 from the G-type form a discrete cluster, as does UGT73C12 from the P-type and UGT73C13 from the G-type. UGTs in the first cluster are more than 95% identical to each other, and those in the second cluster are more than 97% identical (Supplemental Table S1). Accordingly, UGT73C9/UGT73C10 from the P-type correspond to UGT73C11 from the G-type and UGT73C12 from the P-type corresponds to UGT73C13 from the G-type. In comparison with UGT73C homologs from *A. thaliana*, *Arabidopsis lyrata*, and *Brassica rapa*, the five *B. vulgaris* sequences are most closely related to *A. thaliana* UGT73C5 and UGT73C6 and a UGT73C5 homolog in *A. lyrata*.

The UGTs described in the phylogeny have been exposed to different levels of selection since they diverged, as indicated by the significantly better fit of a...
model with independent $\omega$ (ratio of the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site [dN/dS ratios]) for each branch compared with a single common $\omega$ ratio for all branches ($2\Delta \ln L = 13.9; P < 0.001$). Positive selection among branches was further indicated by the better fit of a model including positive selection (model M3) than a model without $M0$ ($2\Delta \ln L = 304.7; P < 0.001$); 4.3% of the codons were estimated to have been under positive selection. Only the branches leading to UGT73C9, UGT73C10, and UGT7311 showed signs of positive selection; branches leading to UGT73C12 and UGT73C13 as well as A. thaliana, A. lyrata, and B. rapa have $\omega < 1$, showing that these branches are under purifying selection.

All five UGT sequences were mapped to an existing linkage map of B. vulgaris (Kuzina et al., 2011) and found to be located in a region that corresponds to A. thaliana chromosome 2 between 13.5 and 19.6 Mb. None of the UGTs lie within previously reported regions containing quantitative trait loci (QTL) for resistance toward P. nemorum larvae feeding (Kuzina et al., 2011). In A. thaliana, six out of the seven UGT3C genes are positioned in a tandem repeat cluster at 15.4 Mb on chromosome 2. Therefore, it is likely that the identified B. vulgaris UGT3C genes are located in a similar UGT3C cluster in the B. vulgaris genome.

Heterologous Expression and In Vitro Activities of the UGT73Cs

To determine if the five UGTs isolated from G- and P-type B. vulgaris have similar catalytic activities as BvUGT1 from B. vulgaris var variegata, they were heterologously expressed in Escherichia coli. The corresponding crude protein extracts were assayed with different sapogenins as putative sugar acceptors and UDP-Glc as sugar donor. UGT73C10, UGT73C11, UGT73C12, and UGT73C13 catalyzed transfer of a Glc moiety from UDP-Glc to the oleanane sapogenins oleanolic acid and hederagenin and to the lupane sapogenin betulinic acid (Fig. 4). In addition, their precursors $\beta$-amyrin and lupeol were glucosylated, but with lower efficiency (Fig. 5). In contrast, UGT73C9 from the P-type appeared inactive toward the compounds tested.

The glucosylation positions of the two oleanane sapogenins produced by the UGTs were determined by NMR spectroscopy. Based on one-dimensional (1-D) $^1$H- and $^{13}$C- as well as two-dimensional (2-D) Correlation Spectroscopy (COSY)-, Total Correlation Spectroscopy (TOCSY)-, and Heteronuclear Single Quantum Coherence (HSQC)-NMR analyses (Supplemental Data Set S1), the glucosides were concluded to be 3-O-$\beta$-D-glucopyranosyl oleanolic acid and 3-O-$\beta$-D-glucopyranosyl hederagenin. This is in agreement with these monoglucosides as predicted precursors of oleanolic acid cellobioside and hederagenin cellobioside, respectively.

In addition to the 3-O-monoglucosides, UGT73C12 and UGT73C13 also formed low amounts of diglucosides, while this activity was barely detectable for UGT73C10 and UGT73C11. Based on retention times and fragmentation patterns in liquid chromatography-mass spectrometry analyses, these diglucosides could not be oleanolic acid and hederagenin cellobioside, respectively, but represent bidesmosidic glucosylation (i.e. glycosylation at two different positions; Supplemental Fig. S1). A diglucosylated betulinic acid was, in addition to two different betulinic acid monoglucosides, produced in detectable amounts after 30 min of incubation when using betulinic acid concentrations as low as 10 $\mu$M (Fig. 5). After alkaline hydrolysis (saponification), which cleaves the ester but not the ether bonds in

---

**Figure 4.** Activity of the heterologously expressed B. vulgaris UGT73Cs toward sapogenins. Enzyme assays contained 750 ng of recombinant UGT in 50 $\mu$L and 50 $\mu$L oleanolic acid (oa), hederagenin (he), or betulinic acid (be) as acceptor substrates and 1 mM UDP-Glc as donor substrate. The assays were incubated for 60 min at 30°C and analyzed by TLC. Compounds were visualized by spraying with 10% sulfuric acid in methanol and subsequent heating. The (inverted) image was taken at long-wave UV (366 nm) excitation. Migration of authentic oleanolic acid, hederagenin, 3-O-$\beta$-Glc oleanolic acid (oa-Glc), and 3-O-$\beta$-Glc hederagenin (he-Glc) is shown in the reference lane (Ref.). Positions of aglycones (aglc), monoglucosides (m-Glc), and diglucosides (di-Glc) are indicated on the left side.
glucosylated products, the betulinic acid diglucoside and one of the two betulinic acid monoglucosides were no longer detectable (Supplemental Fig. S2). Therefore, the degraded monoglucoside must be 28-O-glucosylated betulinic acid and the diglucoside must be 3,28-O-diglucosylated betulinic acid. Similarly, the diglucosidic forms of oleanolic acid and hederagenin would represent 3,28-O-diglucosides. Under assay conditions with high amounts of enzyme, increased incubation time, and elevated incubation temperature, UGT73C13 also produced an oleanolic acid triglucoside (Supplemental Fig. S1), which further demonstrated the lower substrate specificity and regiospecificity of UGT73C13. However, the low in vitro production of these glucosides suggests that these additional activities only play a minor role, if any, in planta.

Other members of the UGT73C subfamily have been assigned to be involved in flavonoid and brassinosteroid metabolism (Jones et al., 2003; Poppenberger et al., 2005; Modolo et al., 2007). Glycosylated flavonoids derived from quercetin and kaempferol are present in B. vulgaris (Senatore et al., 2000; Dalby-Brown et al., 2011). Consequently, the flavonols quercetin and kaempferol, the phyto steroids obtusifoliol, campesterol, sitosterol, and stigmasterol, and the brassinosteroid 24-epi-brassinolide were tested as substrates. 2,4,5-Trichlorophenol (TCP) was included as a positive control, as it can be glycosylated by several different plant UGTs (Messner et al., 2003; Brazier-Hicks and Edwards, 2005). Of the compounds tested, UGT73C9 only showed weak activity toward TCP when applied in 1 mM concentration. In contrast, UGT73C10, UGT73C11, UGT73C12, and UGT73C13 glucosylated TCP at 10 μM concentration (Fig. 5). The levels of oleanolic acid, hederagenin, and betulinic acid glucosides produced by these four UGTs were constantly higher than the levels of TCP glucosides, showing that sapogenins are better substrates. UGT73C10 and UGT73C11 showed weak activity toward quercetin and kaempferol at 100 μM concentration, while at 10 μM, glucosides could not be detected. In contrast, UGT73C12 and UGT73C13 clearly produced flavonol glucosides in assays with 100 μM quercetin or kaempferol, while at 10 μM, the glucosides were hardly detectable (Fig. 5). 24-Epi-brassinolide glucoside(s) were not observed with UGT73C11, whereas UGT73C13 catalyzed glucosylation of 24-epi-brassinolide glucoside, produced by A. thaliana UGT73C5 (Supplemental Fig. S3). None of the B. vulgaris UGTs glucosylated the phytosterols. A. thaliana UGT73B85 was included to represent a UGT73 from a different subfamily than UGT73C. UGT73B85 glucosylated TCP but neither of the sapogenins or other compounds tested (Supplemental Figs. S3 and S13).

UDP-Gal and UDP-GlcA were tested as alternative sugar donors. No glucuronides could be detected with any of the B. vulgaris UGTs when UDP-GlcA was used as sugar donor, but low activity was observed for UDP-Gal (Supplemental Fig. S4). 1H-NMR analysis revealed that the UDP-Gal stock contained traces of UDP-Glc, suggesting that the activity observed most likely originates from the UDP-Glc contamination (Thorsøe et al., 2005).

In summary, UGT73C10, UGT73C11, UGT73C12, and UGT73C13 preferentially glucosylate different oleanane and lupane sapogenins. Both UGT73C10 and UGT73C11 show high regiospecificity and substrate specificity by predominantly glucosylating the C3-hydroxyl group of sapogenins via an ether linkage. In comparison, UGT73C12 and UGT73C13 show lower substrate specificity and also glucosylate the sapogenin C28-carboxyl group via an ester bond. However, the ability to glucosylate at the C28-carboxyl group varied strongly: C28 glucosylation was abundant for betulinic acid and to a lesser extent for oleanolic acid and weakly for hederagenin. The similar enzymatic characteristics of UGT73C10 from the P-type and UGT73C11 from the G-type corroborate the phylogenetic reconstruction (Fig. 3), as do the characteristics of UGT73C12 from the P-type and UGT73C13 from the G-type. UGT73C9 apparently does not glucosylate any of the tested compounds besides the positive control substrate TCP, despite clustering with UGT73C10 and UGT73C11.

Kinetic Parameters of UGT73C11 and UGT73C13

Enzymes in the biosynthesis of plant specialized metabolism are generally characterized by low \( K_m \) and high turnover rates. To evaluate the affinity and catalytic efficiencies of the two UGT clusters (Fig. 3), the

---

**Figure 5.** Substrate specificity of UGT73C10 and UGT73C12. TLC analyses of activity assays with recombinant UGT73C10 or UGT73C12 using \(^{14}\)C-labeled UDP-Glc as donor substrate are shown. Substrates tested were oleanolic acid (oa), hederagenin* (he), \( \beta \)-amyrin (\( \beta \)a), betulinic acid (be), kaempferol (ka), quercetin (qu), and TCP, applied at either 100 or 10 μM concentration. *The hederagenin batch contained a low amount of oleanolic acid.
The kinetic parameters of UGT73C11 and UGT73C13 (both from the G-type) were determined toward hederagenin and oleanolic acid (Table I). Optimal assay conditions were at pH 8.6 for UGT73C11 and pH 7.9 for UGT73C13, with 1 mM dithiothreitol (DTT) as reductant. Purification of the recombinant UGTs was omitted due to decreasing specific activity upon metal chelate affinity-based purification. Instead, recombinant UGT amounts were quantified directly in crude E. coli protein extracts by taking advantage of an introduced N-terminal fused S-tag.

Most of the saturation curves (Supplemental Fig. S6) were hyperbolic and could be described by the Michaelis-Menten equations (for estimates, see Table I). However, for UGT73C13, the reaction velocities decreased when oleanolic acid concentrations exceeded 50 μM, indicating that it inhibits enzyme activity beyond this concentration. Similar substrate inhibition has previously been reported for other family 1 UDP-glycosyltransferases (Luukkanen et al., 2005; Ono et al., 2010). UGT73C11 has a 7-fold lower $K_m$ value and a 3-fold higher turnover rate ($k_{cat}$ value) with hederagenin than UGT73C13. The two UGTs have comparable $K_m$ values with oleanolic acid, but UGT73C11 has a 3.5-fold higher $k_{cat}$ value. The kinetic parameters, therefore, corroborate that UGT73C10 and UGT73C11 have higher affinity for sapogenins and more efficiently catalyze 3-O-glycosylation of oleanolic acid and hederagenin than UGT73C12 and UGT73C13. The low $K_m$ (less than 10 μM) and high $k_{cat}$ values of UGT73C11 are in comparable ranges to flavonol UGTs with their in planta acceptor substrates (Noguchi et al., 2007; Ono et al., 2010). The 1.4-fold higher catalytic efficiency ($k_{cat}/K_m$) for hederagenin than for oleanolic acid indicates that hederagenin is the preferred substrate for UGT73C11. Interestingly, UGT73C13 shows opposite substrate preference, as it has a 3-fold higher $k_{cat}/K_m$ value for oleanolic acid than for hederagenin. The $K_m$ for UDP-Glc was estimated to be around 95 μM for UGT73C11 and 25 μM for UGT73C12 (Supplemental Fig. S5).

### In Vitro Activities of the UGT73Cs toward *B. vulgaris* Sapogenin Mixtures

The saponin composition of *B. vulgaris* is complex, with more than 40 putative saponins detected in liquid chromatography-mass spectrometry analyses (Supplemental Figs. S7 and S8). The majority of these appear specific for either one of the two plant types, while others are present in variable amounts in both types. To evaluate if the UGTs can glucosylate other *B. vulgaris* sapogenins than oleanolic acid and hederagenin, crude saponin-containing extracts of both plant types were subjected to acidic hydrolysis to O-deglycosylate the saponins. Tandem mass spectrometry to n-fold (MSn) fragmentation analyses showed that the saccharide side chains of saponins in both *B. vulgaris* types consist of one to four hexosyl moieties, as concluded from the sequential loss of fragments with a mass of 162 D. The MSn fragmentation patterns of the most intense putative sapogenins in the G-type extract further indicate that they are derived from sapogenins with masses of 456 and 472 D, corresponding to oleanolic acid and hederagenin, as well as 458 and 488 D. In addition, a few less intense putative sapogenins appear to be derived from sapogenins with masses of 470, 474, and 476 D. In metabolite extracts of the P-type, the most abundant putative sapogenins originate from sapogenins with a mass of 474 D, followed by sapogenins derived from 458- and 488-D sapogenins. Only a few putative sapogenins based on sapogenins with masses of 456 and 472 D occur in this plant type.

After acid hydrolyzation, the putative sapogenins could not be detected, which confirms complete deglycosylation (Supplemental Figs. S9 and S10). The hydrolyzed G-type extract contained at least 40 structurally distinct compounds that are likely to be sapogenins, while in the P-type extract, 13 putative sapogenins were detected. Incubation of these extracts with UGT73C10, UGT73C11, UGT73C12, and UGT73C13 and UDP-Glc as sugar donor yielded numerous compounds that, based on MSn fragmentation patterns, were putative sapogenin monoglucosides (Supplemental Fig. S11). For both the G- and P-type sapogenin extracts, incubation with UGT73C10 and UGT73C11 reduced peak intensities of all putative sapogenins and resulted in the formation of the corresponding monoglucosides. In contrast, UGT73C12 and UGT73C13 appeared restricted to glucosylate only a subset of the putative sapogenins. Moreover, monoglucosides were produced at lower rates by UGT73C12 and UGT73C13 compared with UGT73C10 and UGT73C11. As expected, 3-O-β-D-Glc hederagenin (compound G27 in Supplemental Fig. S11) and 3-O-β-D-Glc oleanolic acid

### Table I. Kinetic parameters of UGT73C11 and UGT73C13 toward oleanolic acid and hederagenin

<table>
<thead>
<tr>
<th>UGT</th>
<th>Sapogenin</th>
<th>$K_m$ (μM)</th>
<th>$k_{cat}$ (s⁻¹)</th>
<th>$k_{cat}/K_m$ (s⁻¹ μM⁻¹)</th>
<th>$K_i$ (μM)</th>
<th>$V_{max}$ (nmol min⁻¹ mg⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UGT73C11</td>
<td>Oleanolic acid</td>
<td>9.7 ± 2.2</td>
<td>0.816</td>
<td>0.084</td>
<td>817 ± 118</td>
<td>390 ± 38</td>
</tr>
<tr>
<td></td>
<td>Hederagenin</td>
<td>3.3 ± 0.8</td>
<td>0.389</td>
<td>0.118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UGT73C13</td>
<td>Oleanolic acid</td>
<td>12.5 ± 2.1</td>
<td>0.231</td>
<td>0.19</td>
<td>262</td>
<td>231 ± 21</td>
</tr>
<tr>
<td></td>
<td>Oleanolic acid²</td>
<td>7.6 ± 1.2</td>
<td>0.176</td>
<td>0.23</td>
<td>176 ± 7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hederagenin</td>
<td>22.9 ± 4.8</td>
<td>0.131</td>
<td>0.006</td>
<td>131 ± 10</td>
<td></td>
</tr>
</tbody>
</table>

* Kinetic parameters based on fit to the substrate inhibition equation. ²Kinetic parameters based on fit to the Michaelis-Menten equation.
(compound G35 in Supplemental Fig. S11) were among the products formed from the G-type extract by UGT73C10 and UGT73C11. Surprisingly, only trace amounts of these two sapogenin monoglucosides were observed upon incubation of the G-type extract with UGT73C12 and UGT73C13. These UGTs additionally produced low amounts of diglucosides and compounds that may be kaempferol glucosides (according to their MS$^2$ fragmentation patterns). These findings corroborate that UGT73C12 and UGT73C13 have lower substrate specificity toward sapogenins than UGT73C10 and UGT73C11, which was also concluded from the in vitro enzyme assays (Fig. 5).

**In Planta Saponin Accumulation Correlates with Organ-Specific Expression of the UGT73Cs**

Steady-state transcript levels of the UGT73Cs were determined in leaves, petioles, and roots of 2-month-old G- and P-type *B. vulgaris* plants and compared with saponin accumulation in these organs. Metabolite extracts were evaluated by liquid chromatography-mass spectrometry and revealed a characteristic organ-specific saponin relative abundance in both plant types. Relative accumulation was highest in leaves, intermediate in petioles, and widely absent in roots (Fig. 6A; Supplemental Fig. S12). This pattern was consistent across the different plants tested.

Two primer sets were used to quantify steady-state transcription levels of the UGTs by quantitative real-time PCR (Fig. 6, B and C). Due to the high sequence identities between UGT73C11 in the G-type and UGT73C10 and UGT73C9 in the P-type, it was not possible to design a primer that could differentiate between these three genes. Accordingly, primer set 1 amplifies UGT73C11 in the G-type, while in the P-type it amplifies simultaneously UGT73C9 and UGT73C10. Similarly, primer set 2 amplifies UGT73C13 from the G-type and UGT73C12 from the P-type. All plants showed the highest expression of UGT73C11 and UGT73C9/C10 in leaves, an up to 10-fold lower expression in petioles, and up to 200-fold lower expression in roots, despite some variation among individual plants tested. A similar expression pattern was observed for UGT73C13 and UGT73C12. In general, UGT73C11 and UGT73C9/C10 were expressed at a higher level than UGT73C13 and UGT73C12. The highest expression level of UGT73C13 was observed in plants with the lowest UGT73C11 expression. Since those plants were in a more progressed developmental stage (Supplemental Fig. S12), this suggests alternating expression regulation of the two genes during plant ontogenesis.

**3-O-β-D-Glc Hederagenin Is a Feeding Deterrent against *P. nemorum***

The two diglucosides hederagenin and oleanolic acid cellobioside have previously been shown to deter feeding by *P. nemorum* (Nielsen et al., 2010). To determine if the corresponding monoglucosides have a similar effect, approximately 12.5 mg of 3-O-β-D-Glc hederagenin and 8.5 mg of 3-O-β-D-Glc oleanolic acid were produced in vitro with UGT73C10 (see above).

![Figure 6. Comparison of relative saponin abundance and expression of the UGTs in different *B. vulgaris* organs. A, Relative saponin abundance in leaf, petiole, and root extracts of three G-type plants (G1–G3), based on the mean peak areas ± SD of the extracted ion chromatograms from liquid chromatography-mass spectrometry of the four insect resistance-correlated G-type saponins: hederagenin cellobioside (he-cell), oleanolic acid cellobioside (oa-cell), gypsogenin cellobioside (gy-cell), and 4-epi-hederagenin cellobioside (4e-cell).Overlayed base peak chromatograms of all liquid chromatography-mass spectrometry runs are provided in Supplemental Figure S12. B, Expression of UGT73C11 in the three G-type plants (G1–G3) and combined expression of UGT73C9 and UGT73C10 in three P-type plants (P1–P3), determined with primer set 1 relative to actin (ACT2). Values are means ± SD of technical duplicates. C, Corresponding expression analysis of UGT73C13 in G1 to G3 and UGT73C12 in P1 to P3, determined with primer set 2.](image-url)
Both compounds were painted on 92-mm² radish (Raphanus sativus) leaf discs in doses of 3.75, 15, and 60 nmol and presented to P. nemorum adults of either the susceptible (ST; rr genotype) or resistant (AK; Rr genotype) line, and the area consumed was evaluated after 24 h (Fig. 1).

3-O-β-D-Glc hederagenin significantly reduced the leaf consumption by susceptible ST beetles, with dose-dependent reductions of 26%, 55%, and 92% in response to 3.75, 15, and 60 nmol per leaf disc, respectively (Fig. 7A; the reduction by 15 and 60 nmol was statistically significant [P < 0.005] when tested separately). A dose-dependent reduction of leaf consumption was also observed for the resistant AK line, with 16% and 67% reduction in response to 15 and 60 nmol, respectively (only the reduction by 60 nmol was significant when tested separately).

3-O-β-D-Glc oleanolic acid had a significantly weaker effect on leaf consumption for both P. nemorum lines (Fig. 7B). Only the high dose of 60 nmol reduced consumption by the ST line (45% reduction), whereas there was no effect on the AK line at any dose. Feeding assays with 3.75 nmol were not conducted, as there was no significant effect with 15 nmol.

When tested in a joint linear mixed-effect model, there was a significant three-way interaction between sapogenin monoglucosides, their doses, and the P. nemorum lines, with a significance level of $P < 0.0001$. Thus, (1) 3-O-β-D-Glc hederagenin is more effective than 3-O-β-D-Glc oleanolic acid, (2) the feeding deterrence of the sapogenin monoglucosides is dose dependent, and (3) the efficacy toward the susceptible P. nemorum line is higher than toward the resistant line.

**DISCUSSION**

Saponin biosynthesis is not fully understood, nor is the relationship between the different chemical structures and their roles in plant defense. Here, we have identified two UGTs that specifically glucosylate sapogenins in the wild crucifer B. vulgaris. These UGTs have evolved to be specific for 3-O-glucosylation of sapogenins. Previously, UGTs that glucosylate sapogenins at the C28 carboxylic groups have been identified in Medicago truncatula (UGT73F3; Naoumkina et al., 2010) and in Saponaria vaccaria (UGT74M1; Meesapyodsuk et al., 2007). Monoglucosylated 3-O-β-D-Glc hederagenin, produced in vitro by one of the UGTs identified here, UGT73C10, is a strong feeding deterrent against P. nemorum, demonstrating that 3-O-glucosylation of saponins is essential for bioactivity. The UGTs are expressed in both a P. nemorum resistant and a susceptible type of B. vulgaris, which fits our observation that most, if not all, saponins in the P and G-types are 3-O-glucosylated. The presence of UGTs in both the plant types catalyzing 3-O-glucosylation sapogenins, and the genomic locations of genes coding for these UGTs outside QTL associated with resistance to P. nemorum, suggest that the difference in resistance between the two B. vulgaris types is determined by an earlier enzymatic step in saponin biosynthesis.

**UGT73C10/C11: Two Neofunctionalized UDP-Glc: Sapogenin 3-O-Glucosyltransferases**

Of the five UGTs we identified in B. vulgaris ssp. arvensis, UGT73C10 from the insect-susceptible P-type and UGT73C11 from the resistant G-type showed highest activity and specificity toward a wide range of sapogenins. Both enzymes exhibit high regiospecificity by preferably glucosylating the C3 hydroxyl group, which is in agreement with structures of saponins in both B. vulgaris types. Both enzymes, in contrast, were essentially inactive toward the flavonols and phytosterols tested. Their acceptor substrate specificity thus differs substantially from other characterized members of the UGT73C subfamily. UGT73C8 from M. truncatula glucosylates several (iso)flavanoids in vitro (Modolo et al., 2007). A. italicana UGT73C6 was suggested to be a UDP-Glc: flavonol-3-O-glycoside-7-O-glucosyltransferase by Jones et al. (2003), based on in vitro activities and T-DNA knockout lines. Recent studies show that UGT73C6 is...
functionally similar to the well-studied UGT73C5, also from *A. thaliana*, in its ability to glucosylate brassinosteroids in overexpression lines (Husar et al., 2011). UGT73C5 in addition glucosylates numerous structurally diverse acceptor substrates (Lim et al., 2003, 2004; Poppenberger et al., 2003, 2005, 2006; Hou et al., 2004; Weis et al., 2006; Caputi et al., 2008). It was originally identified as a mycotoxin-detoxifying enzyme (Poppenberger et al., 2003), but recently, it was suggested to be involved in brassinosteroid homeostasis (Poppenberger et al., 2005). In our study, *A. thaliana* UGT73C5 also glucosylated oleanolic acid, hederagenin, and betulinic acid in vitro, providing further evidence for the promiscuity of this enzyme (Supplemental Fig. S13).

However, it had substantially lower catalytic efficiency and regiospecificity toward oleanolic acid and hederagenin than UGT73C11 and UGT73C13 from *B. vulgaris* (Supplemental Fig. S13). *A. thaliana* is not known to produce triterpenoid saponins or sapogenins, although triterpenoids such as β-amarin and lupeol accumulate in cuticular waxes of stems, siliques, and buds (Shan et al., 2008). Therefore, it is unlikely that the in vitro activities of UGT73C5 with sapogenins reflect an in planta function.

The broad substrate affinity commonly found for some UGTs has been proposed to enable flexibility in response to changes in metabolite profiles (Vogt and Jones, 2000). Specialized enzymes for new biosynthetic pathways may originate from broad progenitor enzymes and are generally characterized by having a lower $K_m$ (and thus higher substrate specificity) and higher catalytic efficiency ($k_{cat}/K_m$) than their more promiscuous progenitors (Jensen, 1976; Aharoni et al., 2005; Ktherosinsky and Tawfik, 2010). Ancestors of UGT73C10/C11 from *B. vulgaris* could thus have been promiscuous UGT73C5-like enzymes that evolved a more narrow specificity and higher efficiency for catalyzing sapogenin 3-O-glucosylation. Based on our analyses, UGT73C12/C13 have broader substrate and product specificities and could represent evolutionary intermediates to UGT73C10/C11 or UGTs specialized in glucosylation of yet unknown sapogenins in *B. vulgaris*.

Our phylogenetic reconstruction shows that the five *B. vulgaris* UGT73Cs indeed cluster separately from the UGT73Cs in *A. thaliana, A. lyrata*, and *B. rapa* (Fig. 3). It further suggests that UGT73C10, UGT73C11, and UGT73C9 originate from a gene duplication event after the split from *A. thaliana* and *B. rapa* and before the P and G-types separated. Another gene duplication separated UGT73C9 from UGT73C10, probably in the P-type after the P- and G-types split. Alternatively, this duplication occurred before the P-G bifurcation and the UGT73C9 copy was lost subsequently in the G-type.

Of the UGTs in our phylogenetic analysis, UGT73C9, UGT73C10, and UGT73C11 showed clear signs of positive selection during their differentiation. This corroborates our biochemical data, which show that UGT73C10 and UGT73C11 have evolved to a new specialized function. In contrast, UGT73C12 and UGT73C13 showed no signs of selection, corroborating that they have not evolved new biochemical functions; this further suggests that they may be orthologs of *A. thaliana* UGT73C5 or UGT73C6. The observation that UGT73C9 is under positive selection questions the function of this UGT in saponin biosynthesis. Based on our biochemical data, UGT73C9 appears as an expressed pseudogene; however, the phylogenetic analysis indicates that the gene has been under positive selection. An alternative hypothesis is that the substrate for UGT73C9 was not included in our analysis. As the saponin profiles of P- and G-type *B. vulgaris* differ, UGT73C9 could possibly be involved in the differentiation of these.

Genes for the *B. vulgaris* UGTs were located in a genomic region syntenic to a part of *A. thaliana* chromosome 2, which contains a tandem repeat cluster of UGT73Cs. Our recent genome sequencing indicates that the *B. vulgaris* UGT73Cs identified here are also part of a repetitive cluster containing several UGT-like repeats and in higher number than the corresponding UGT73C cluster in *A. thaliana*. This supports that UGT73C10/C11 evolved via gene duplications from a broad-spectrum UGT73C in a common ancestor shared with *A. thaliana*, as discussed above. It further supports the idea that the evolution of novel bioactive metabolites often occurs via gene duplication and neofunctionalization (Osbourn, 2010; Weng et al., 2012) followed by increased specialization (Jensen, 1976; Aharoni et al., 2005; Ktherosinsky and Tawfik, 2010).

3-O-Glucosylation of Hederagenin Deters Feeding by *P. nemorum*

Monoglucosylation of hederagenin into 3-O-β-D-Glc hederagenin clearly suppressed feeding by *P. nemorum*. A similar but lower suppression was found for 3-O-β-D-Glc oleanolic acid. The diglucosylated forms of hederagenin and oleanolic acid (hederagenin cellobioside and oleanolic acid cellobioside) have previously been found to suppress feeding (Nielsen et al., 2010), in contrast to the aglycones (hederagenin and oleanolic acid). Our results now show that glucosylation with only a single glucosyl group is enough to affect herbivores. The amount of monoglucosides used in our feeding assays was comparable to natural levels of hederagenin cellobioside in *B. vulgaris* leaves (Shinoda et al., 2002), and our results thus demonstrate that 3-O-β-D-Glc hederagenin and 3-O-β-D-Glc oleanolic acid are biologically relevant feeding deterrents. Furthermore, the higher efficiency of hederagenin than oleanolic acid, in both their monoglucosylated and diglucosylated forms, shows that C23 hydroxylation in the hederagenin backbone increases this antifeedant effect.

The precise mechanism that enables glucosylated sapogenins to deter insects is not known. The dependency on glycosylation indicates that membrane perturbation plays a role, at least for *P. nemorum*. In agreement with this, saponins have been shown to damage the midgut epithelium of pea aphids (*Acyrthosiphon pisum*; De Geyer et al., 2012). Alternatively, glucosylated saponins may have a more adverse taste for insects than the corresponding sapogenins (Glendinning, 2002); however, *P. nemorum* larvae die from exposure to G-type leaves (Nielsen, 1997a).
Evolution of Saponin Biosynthesis in *Barbarea* Species

The multitude of different putative sapogenins in the G- and P-types indicates that OSCs and P450s are responsible for much of the saponin diversity in this species and probably for the differences between the two plant types. The phylogeny of OSCs (Phillips et al., 2006; Augustin et al., 2011) suggests frequent changes in product spectra during evolution, which is supported by the drastic spectrum changes that may arise from only a few amino acid substitutions (Lodeiro et al., 2005). Changes in cytochrome P450 activity are also known to affect saponin profiles and activity. Carelli et al. (2011) showed that lack of a functional CYP716A12, which catalyzes C28 carboxylation of triterpenoid sapogenins, results in a complete loss of hemolytic sapogenins in *M. truncatula*. In contrast, nonhemolytic saponins were unaffected. The nonhemolytic saponins are derived from sapogenins that are not carboxylated at the C28 position, and MS² fragmentation of these revealed an aglycone fragment ion with a deduced mass of 474 D (Pollier et al., 2011). A similar fragmentation product was observed for P-type saponins and suggests that structurally similar saponins, with four hydroxyl groups but no C28 carboxylation, are present in this plant type. Different abilities to catalyze C28 oxygenation by cytochromes P450 could thus be involved in determining the different structures of G- and P-type saponins and thus their effect on insect herbivores.

The current hypothesis for the evolution of insect resistance in *B. vulgaris* suggests that it took place after the first species of the *Barbarea* genus had emerged (Agerbirk et al., 2003b; the age of this split is unknown at present). An OSC probably mutated to be able to catalyze the conversion of oidosqualene into saponin precursors, which is in agreement with the presence of triterpenoids in *A. thaliana*. Later, UGTs must have evolved to become specific to the novel sapogenins produced by the resistant *Barbarea* species, as we have shown here. Whether the cytochromes P450 involved in saponin biosynthesis of *Barbarea* species have also specialized is not known. Much later, *B. vulgaris* differentiated into the G- and P-types, possibly during one of the last ice ages (Hauser et al., 2012; Toneatto et al., 2012). Thus, the two plant types are genetically and geographically differentiated, reproductively somewhat incompatible, and differ for several traits apart from insect resistance and saponin structure (Toneatto et al., 2010; Dalby-Brown et al., 2011). Thus, the most likely scenario suggests that the P-type lost resistance to *P. nemorum* during this allopatric separation. Our results here clearly show that this loss of insect resistance was not caused by a loss of UGT function. Instead, we have shown that UGTs of *B. vulgaris* have adapted to the evolutionary gain of sapogenins in this species.

**MATERIALS AND METHODS**

**Activity-Based cDNA Library Screening**

*Barbarea vulgaris var. variegata* (Chiltern Seeds) leaf RNA was used for first-strand synthesis with the ZAP-cDNA Synthesis Kit (Stratagene). The resulting cDNA was digested with XhoI, ligated into the predigested Uni-ZAP XR vector (Stratagene), and transformed into the Escherichia coli strain XL1-Blue MRF² (Stratagene). After in vivo excision of pBluescript SK⁻ phagemids from the Uni-ZAP XR vectors, the obtained *E. coli* colonies were combined in terrific broth (TB) medium and transferred to 96-well plates (approximately 100 colonies per well). The *E. coli* suspensions were incubated with shaking at 37°C for 3 h and then for 3 h with 0.1 mM isopropylthio-β-galactoside (IPTG). Cultures of individual wells were combined into batches (four wells per batch), and the bacterial cells were harvested by centrifugation and lysed by sonication. Enzymatic activity was tested by incubating the lysates overnight at 30°C with 200 μM UDP-Glc and 175 μM oleanolic acid. Ethyl acetate extracts of the activity assays were analyzed by TLC on Silica Gel 60 F254 plates (5554; Merck), using chloroform:methanol:water (32:9:1) as mobile phase, and stained by spraying with 10% sulfuric acid in methanol followed by heating. Batches that showed oleanolic acid glucosylation activity were in additional screening rounds stepwise further diluted until a single active clone designated BvUGT1 was identified.

**Cloning of BvUGT1 Homologs from *B. vulgaris* ssp. arcuata**

Contigs representing fragments of BvUGT1 homologs were identified in a 454 pyrosequencing-generated transcriptomic G-type data set (Kuzina et al.,...
2011) using local BLASTx. Total RNA was extracted from leaves of G- and P-type B. vulgaris using the NucleoSpin RNA Plant kit (Macherey-Nagel) and 3’ RACE performed with the FirstChoice RLM-RACE kit (Ambion) according to the manufacturer’s protocol. The applied primers are listed in Supplemental Table S2.

The nucleotide sequences of UGT73C9, UGT73C10, UGT73C11, UGT73C12, and UGT73C13 were cloned from genomic DNA of an F1 hybrid plant, which originated from crossings between G- and P-type plants (Kuzina et al., 2009), and ligated into pGEM-T Easy for sequencing. PCRs for cloning were performed with Phusion High-Fidelity DNA Polymerase (Finnzymes), and PCRs for screening and A-tailing reactions were performed with HotMaster Taq DNA Polymerase (Sprint). A-tailing reactions were set up according to the pGEM-T Easy manual (Promega). Sequencing was performed by Eurofins MWG Operon.

Phylogenetic Analysis

UGT73 amino acid sequences were aligned (Supplemental Data Set S2) using MUSCLE and used to construct a maximum likelihood bootstrapped phylogenetic tree using MEGA (version 5.0; Jones, Taylor, and Thornton substitution model, uniform rates among sites, 100 bootstrap replications; Tamura et al., 2011). The A. thaliana hys and Brassica napus UGTs, identified by BLAST searches at www.phytozome.net and www.brassica-rapa.org, have not been officially named and are therefore named here according to their grouping with Arabidopsis thaliana.

To test for signs of past selection on the UGTs, branch and site models were estimated using code in the PAML package (http://abacus.gen.ucl.ac.uk/software/pamlhtm). For positive selection between branches, the free-ratio model was compared with the one-ratio model and tested by comparing the twice log-likelihood difference between models to an $\chi^2$ distribution with 18 degrees of freedom. Seven site models were estimated: M0 (one ratio); M1 (nearly neutral; two categories); M2 (positive selection; three categories); M3 (discrete; three categories); M5 (5 γ; 10 categories); M7 (β; 10 categories); and M8 ($β_{ω > 1}; 1$ category); these were tested as above with degrees of freedom corresponding to the number of parameters for the models tested.

Locating UGT73C9, UGT73C10, UGT73C11, UGT73C12, and UGT73C13 on the B. vulgaris Linkage Map

The five UGTs were mapped using the derived cleaved amplified polymorphic sequences or cleaved amplified polymorphic sequences technique. PCR was performed using genomic DNA of an F2 segregating population generated from a cross between P- and G-type B. vulgaris (Kuzina et al., 2009). PCR products obtained using primers eFPsior and seperry (UGT73C9 to -C1), eFPsior and seperry (UGT73C12/13), or Inf and dCapsAvAI (UGT73C11) were digested with EcoRI, BamHI, AxlII, or PvuI to discriminate between UGT73C9, UGT73C10, UGT73C11, and UGT73C13, respectively. Data were scored and analyzed as described by Kuzina et al. (2011).

Heterologous Expression of B. vulgaris UGT73Cs

N-terminally His-tagged expression constructs of UGT73C9, UGT73C10, UGT73C11, UGT73C12, and UGT73C13 were obtained by subcloning into the Ndel and BamHI restriction sites of the pET28c vector (Novagen). N-terminally S-tag expression constructs of the five UGT73 ORFs were achieved by Gateway cloning into pJAM1786 (Luo et al., 2007). For heterologous expression of the His-tag and S-tag constructs, expression vectors were transformed into the E. coli strain Xfbi(DES) (Zymo Research). Expression was carried out in 25-mL Erlenmeyer flasks and started by inoculating 2 mL of Luria-Bertani medium, containing either 50 μg mL$^{-1}$ kanamycin (His-tag constructs) or 100 μg mL$^{-1}$ carbenicillin (S-tag constructs), with a single colony. A 12-h incubation phase at 30°C and 220 rpm was followed by the addition of 4 mL of TB medium containing appropriate selection antibiotics. Ara and IPTG were added to final concentrations of 3 and 0.1 mM, respectively, and the cultures were incubated for 24 h at 15°C and 220 rpm. For expression of the S-tag constructs, 1 μL of 50 mg mL$^{-1}$ carbenicillin mL$^{-1}$ culture was added approximately 12 h after the addition of the TB medium. Bacteria were harvested in aliquots corresponding to 2 mL of culture with an optical density of 8.0, resuspended in 750 μL aliquot $\times 1$ mL HEPES, pH 7.8, and stored at $-80°C$. Bacteria were lysed by thawing aliquots at room temperature. The viscosity of lysates was lowered by incubation with DNase (AppliChem) treatment (1 μg mL$^{-1}$). Cell debris were removed by centrifugation, and supernatants were used as crude protein extracts for enzyme assays. Quantification of heterologously expressed enzymes, fused to an S-tag within E. coli crude protein extracts, was carried out using the FRETWorks S-tag assay kit (Novagen) according to the manufacturer’s protocol.

Substrate Specificity Assays

Enzyme assays to determine substrate specificity were performed in a final volume of 20 μL, containing 2 μL of E. coli crude protein extract with recombinant UGT73C9, UGT73C10, UGT73C11, UGT73C12, or UGT73C13 coupled to a S-tag. Reaction conditions were 25 mM TAPS-HCl, pH 8.6, 1 mM UDP-Glc (Sigma-Aldrich), and 3.31 μM [14C]UDP-Glc (Perkin-Elmer). Ethanol was removed from the UDP-[14C]Glc stock by evaporation prior to setting up the assays. Enzyme assays were started by addition of the acceptor substrates solubilized in dimethyl sulfoxide (DMSO) to final concentrations of 1 mM (only TCP), 100 μM, or 10 μM of the acceptor substrate and 6.25% to 10% (v/v) DMSO, respectively. Reactions were incubated for 30 min at 30°C and stopped by the addition of 130 μL of methanol. Precipitated proteins were removed by centrifugation. Solvent from the supernatant was removed with a vacuum concentrator, and metabolites were dissolved in 20 μL of 50% ethanol and analyzed by TLC. TLC plates were developed in ethyl acetate:methanol:formic acid:water (7.5:0.5:1:1). Radioactive bands were visualized using a STORM 840 PhosphorImager (Molecular Dynamics).

The nucleotide sequences of all UGTs, including 3’-UTRs, were mapped using the derived cleaved ampli
ed polymorphic sequences technique. 14C-labeled monoglucosides were obtained by overnight incubation of oleanolic acid or hederagenin [14C]monoglucoside dilution series. Signal intensities were quantified using ImageQuant 5.0 (Molecular Dynamics). $K_m$ and $V_{max}$ values were calculated using SigmaPlot 11.0 (Systat Software) for nonlinear regression according to the Michaelis-Menten equation or the velocity equation for substrate inhibition.

Determination of Enzyme Kinetic Parameters

Freshly lysed E. coli crude protein extracts were diluted in 10 mM TAPS-HCl, pH 8.0, and 10 mM bovine serum albumin (BSA) to final concentrations of 5 ng μL$^{-1}$ S-tag UGT73C11 and 45 ng μL$^{-1}$ S-tag UGT73C13. The diluted crude protein extracts were applied in master mixtures with final reaction conditions as follows: 25 mM TAPS-HCl, pH 8.6 (UGT73C11) or pH 7.9 (UGT73C13), 1 mM DTT, 500 μM UDP-Glc, 2 mg mL$^{-1}$ BSA, and 0.5 ng μL$^{-1}$ UGT73C11 or 4.5 ng μL$^{-1}$ UGT73C13. Enzyme assays were performed in a volume of 20 μL. Concentrations of UDP-[14C]Glc (Perkin-Elmer) in the total amount of UDP-Glc ranged from 3.31 μM (0.04 kBq μL$^{-1}$ ) to 33.12 μM (0.37 kBq μL$^{-1}$) to ensure sufficient signal intensity. Oleanolic acid and hederagenin were dissolved in 100% DMSO and assayed in duplicate in final concentrations ranging from 0.125 to 8 μM for UGT73C11 and 1.56 to 100 μM for UGT73C13, but with a constant final DMSO concentration of 6.25%. Reactions were preincubated for 5 min at 30°C prior to addition of the acceptor substrate. After incubation for 5 min at 30°C, enzymatic activities were stopped by the addition of 50 μL of ethyl acetate. Assays were extracted four times with 50 μL of ethyl acetate, and the solvent from the combined extractions was removed by evaporation in a vacuum concentrator. Metabolites were dissolved in 96% ethanol and analyzed by TLC. TLC plates were developed using dichloromethane:methanol:water (80:19:1) as mobile phase and visualized as described above. Products were quantified by codeveloping TLC plates with a defined oleic acid or hederagenin [14C]monoglucoside dilution series. Signal intensities were quantified using ImageQuant 5.0 (Molecular Dynamics). $K_m$ and $V_{max}$ values were calculated using SigmaPlot 11.0 (Systat Software) for nonlinear regression according to the Michaelis-Menten equation or the velocity equation for substrate inhibition.

14C-labeled monoglucosides were obtained by overnight incubation of 20 nmol of oleic acid and hederagenin with UGT73C11 at reaction conditions similar to those applied for the actual enzyme assays (500 μM UDP-Glc including 33.12 μM UDP-[14C]Glc [0.37 kBq μL$^{-1}$]). Complete conversion of the aglycones was confirmed by TLC analysis of aliquots of these reactions.

Plant Material

B. vulgaris ssp. arvensis seeds were collected in natural populations in Denmark. G-type (Amager; 55°38’N, 12°34’E) and P-type (Tisvåg; 55°36’N, 11°18’E). Plants were grown at 20°C, 16 h of light/8 h of darkness, and 70% to 75% air humidity, fertilized once a week, and the soil was treated with Bac-timos L (Abbott Laboratories) whenever necessary.
Comparison of Saponin Levels and in Planta Expression of UGT73Cs

To determine saponin levels, metabolites were extracted from 20 to 30 mg of ground, lyophilized leaf, petiole, and root material by boiling for 10 min with 37.5 µL of 55% ethanol per mg of tissue powder. Samples were cooled on ice and centrifuged to remove insoluble particles. Supernatants were kept for more than 2 h at −20°C and centrifuged to remove precipitates. Extracts were filtered (polyvinylidene difluoride; 0.45 µm) and transferred to glass sample vials and kept at −20°C for liquid chromatography-mass spectrometry analysis. An Agilent 1100 Series LC device (Agilent Technologies), equipped with a Gemini NX column (35°C; 2.0 × 150 mm, 35 µm; Phenomenex) and coupled to a Bruker HCT-Ultra ion-trap mass spectrometer (Bruker Daltonics), was used for spectrometric analysis. Mobile phases were eluent A, water with 0.1% (v/v) formic acid, and eluent B, acetonitrile with 0.1% (v/v) formic acid. The gradient program was as follows: 0 to 1 min, isocratic 12%; B to 33 min, linear gradient 12% to 80%; B to 35 min, linear gradient 80% to 99%; B to 38 min, isocratic 99%; B to 45 min, isocratic 12%. B at a constant flow rate of 0.2 mL min⁻¹. The detector was operated in negative electrospray mode and included tandem mass spectrometry to two stages (MS²) and three stages (MS³). Chromatograms were analyzed with DataAnalysis 4.0 (Bruker Daltonics), and saponin abundance was calculated based on summed extracted ion chromatograms of all adduct ions.

RNA was extracted from 100 to 150 mg of ground leaf, petiole, and root material by incubation for 10 min with 900 µL of prewarmed hexadecl-trimethylammonium bromide extraction buffer (Chang et al., 1993) at 65°C and 660 rpm. After 2·fold extraction with 900 µL of chloroform-isoamyl alcohol, RNA was precipitated overnight (4°C) from the supernatant by the addition of LiCl to a final concentration of 2 M. Pellets were dissolved in 500 µL of sodium chloride-Tris-EDTA buffer (lev Provost et al., 2007; prewarmed to 65°C) containing 0.1% SDS. RNA was extracted with chloroform-isoamyl alcohol and precipitated from the aqueous phase by adjusting the NaCl concentration to 0.67 M, adding 1 volume of isopropanol, and subsequent incubation for 5 h at −20°C. RNA pellets were washed with 70% ethanol (−20°C), dried, and redissolved in 30 µL of diethyl pyrocarbonate-treated water. The remaining genomic DNA was removed by on-column DNase treatment using the RNAsesy Mini Kit (Qiagen). RNA extracts were assessed for purity and quantified with a NanoDrop ND-1000 (NanoDrop Technologies) and a 2100 Bioanalyzer (Agilent Technologies).

Reference gene sequences were obtained by mapping the 454 pyrosequencing-derived reads of G- and P-type leaf RNA preparations (V. Kuzina and S. Bak, unpublished data) to a data set consisting of all A. thaliana cDNA sequences (TAIR9_cDNA_20006019) using the CLC Genomics Workbench (CLC bio). Two primer pairs, ACT2_for/ACT2_rev1 and ACT2_for2/ACT2_rev2, were designed from reads mapped to A. thaliana ACT2 (AT3G18780). With the exception of four single-nucleotide polymorphisms in an intron region of the ACT2_for/ACT2_rev1 product from the P-type, sequences derived for each primer set from the two plant types were 100% identical. The sequence identity of the two PCR products to the A. thaliana ACT2 ORF were 91% and 96%, respectively, while the encoded protein sequences were 100% identical to A. thaliana ACT2.

Threshold cycle values of the two primer sets were almost identical of the two PCR products to the A. thaliana ACT2 ORF were 91% and 96%, respectively, while the encoded protein sequences were 100% identical to A. thaliana ACT2. Monoglucosides for NMR and Bioassays

For large-scale production of hederaegenin and oleanolic acid monoglucosides, four 2-L Erlenmeyer flasks, containing 250 mL of TB medium with 50 µg mL⁻¹ kanamycin, were inoculated with fresh Xf(D3) colonies harboring the pET28c-UGT73C10 plasmid and incubated for 12 h at 30°C and 180 rpm. After 2·fold extraction with 900 µL of chloroform-isoamyl alcohol, RNA was extracted with chloroform-isoamyl alcohol and precipitated from the aqueous phase by adjusting the NaCl concentration to 0.67 M, adding 1 volume of isopropanol, and subsequent incubation for 5 h at −20°C. RNA pellets were washed with 70% ethanol (−20°C), dried, and redissolved in 30 µL of diethyl pyrocarbonate-treated water. The remaining genomic DNA was removed by on-column DNase treatment using the RNAsesy Mini Kit (Qiagen). RNA extracts were assessed for purity and quantified with a NanoDrop ND-1000 (NanoDrop Technologies) and a 2100 Bioanalyzer (Agilent Technologies).

Reference gene sequences were obtained by mapping the 454 pyrosequencing-derived reads of G- and P-type leaf RNA preparations (V. Kuzina and S. Bak, unpublished data) to a data set consisting of all A. thaliana cDNA sequences (TAIR9_cDNA_20006019) using the CLC Genomics Workbench (CLC bio). Two primer pairs, ACT2_for/ACT2_rev1 and ACT2_for2/ACT2_rev2, were designed from reads mapped to A. thaliana ACT2 (AT3G18780). With the exception of four single-nucleotide polymorphisms in an intron region of the ACT2_for/ACT2_rev1 product from the P-type, sequences derived for each primer set from the two plant types were 100% identical. The sequence identity of the two PCR products to the A. thaliana ACT2 ORF were 91% and 96%, respectively, while the encoded protein sequences were 100% identical to A. thaliana ACT2. Threshold cycle values of the two primer sets were almost identical in quantitative real-time PCR tests on leaf, petiole, and root tissues from a single G-type plant (±0.08±0.26). In addition, threshold cycle values across the three investigated tissues were found widely constant, with a range of ±0.31.

Five micrograms of RNA from each leaf, petiole, and root extract was applied in 100-µL reactions for cDNA synthesis using the iScript cDNA Synthesis Kit (Bio-Rad) according to the manufacturer’s instructions. Quantitative real-time PCR experiments were performed with the DyNaNo Flash SYBR Green quantitative real-time PCR Kit (Finnzymes) in 20-µL reactions according to the manufacturer’s instructions by adding 1 µL of the cDNA preparations as template per reaction. Primer pairs were RYE_for and RYE_rev (UGT73C9 to −C1), RTII_for and RTII_rev (UGT73C12/C13), as well as ACT2_for1 and ACT2_rev1 (ACT2). Duplexes of each setup were run on a QIagen Rotor-Gene Q Real-Time PCR cycler with settings for melting, annealing, extension, and acquiring of 10 s at 95°C, 10 s at 65°C, 20 s at 72°C, and 1 s at 76°C, respectively. The expression analysis was performed using LinRegPCR (version 12.7; Ramakers et al., 2003; Ruiter et al., 2009). Relative expression values were calculated as the ratios of the starting concentrations (N0) given for the ACT2 reference and the corresponding UGT73C primer sets in the LinRegPCR output.
containing oleanolic acid and hederagenin glucosides were collected and evaporated to dryness using a vacuum concentrator.

The purified hederagenin and oleanolic acid monoglucosides were dissolved in NMR-suitable methanol-d4 (Sigma-Aldrich), and NMR spectra were recorded at room temperature on a Bruker Avance D8X 500-MHz NMR spectrometer (Bruker Daltonics) equipped with a broadband inverse probe. Acquired data were calibrated according to the residual solvent peaks at 3.31 ppm for H3 spectra and 40.01 ppm for 13C spectra. For structural elucidation of the two monoglucosides, 1-D 1H and 13C as well as 2-D COSY, TOCSY, and HSQC experiments were performed and compared with corresponding spectra of oleanolic acid and hederagenin and reported NMR data of structurally related compounds (Supplemental Data Set S1).

**Phyllotreta nemorum Feeding Assays**

Nonchoice feeding assays were performed as described previously by Nielsen et al. (2010). Briefly, purified 3-O-β-D-Glc hederagenin and 3-O-β-D-Glc oleanolic acid were in final concentrations of 2, 0.5, and 0.125 mM dissolved in 75% ethanol. Sapogenin monoglucoside solution (15 μL) was painted on both sides of 95-mm² radish (Raphanus sativus) leaf discs, which resulted in doses of 60 nmol (632 pmol mm⁻²), 15 nmol (158 pmol mm⁻²), and 3.75 nmol (39 pmol mm⁻²) sapogenin monoglucoside per leaf disc. Control leaf discs were treated with solvent only. Two identically treated leaf discs were exposed to one beetle for 24 h. Consumed leaf area was measured with a stereomicroscope. For the origin and maintenance of the two flea beetle (P. nemorum) lines, see Nielsen et al. (2010).

Results were analyzed using the R software package (www.r-project.org). The linear effect model allowed for a possible correlation between measures from the same beetle. The starting model included a three-way interaction between beetle line, compound type, and dose; a 5% significance level was used for saponin abundance and UGT73C9 to -C13 expression correlation analysis.

**Supplemental Figure S1.** Comparison of UGTs from *B. vulgaris* (UGT73C11, UGT73C13, and UGT73C9) and *A. thaliana* (UGT73C5 and UGT73B8) in their activity toward hederagenin, 24-epi-brassinolide, sapogenin 3-O-glucosylation products derived from UGT73C13 activity.

**Supplemental Data**

The following materials are available in the online version of this article.

**Supplemental Figure S1.** Comparison of oleanolic acid glucosylation products after long-term incubation of oleanolic acid with UGT73C10, UGT73C11, UGT73C12, and UGT73C13 with oleanolic acid cellobiose.

**Supplemental Figure S2.** Alkaline hydrolysis (saponification) of betulinic acid glucosylation products derived from UGT73C13 activity.

**Supplemental Figure S3.** Comparison of UGTs from *B. vulgaris* (UGT73C11, UGT73C13, and UGT73C9) and *A. thaliana* (UGT73C5 and UGT73B8) in their activity toward hederagenin, 24-epi-brassinolide, and TCP.

**Supplemental Figure S4.** Comparison of UDP-Glc and UDP-Gal as sugar donor substrates of UGT73C10.

**Supplemental Figure S5.** Determination of *Kₘ* values of UDP-Glc for UGT73C11 and UGT73C12.

**Supplemental Figure S6.** Kinetics of UGT73C11 and UGT73C13 with oleanolic acid and hederagenin as acceptor substrates.

**Supplemental Figure S7.** Liquid chromatography-mass spectrometry analysis of a G-type *B. vulgaris* metabolite extract with 55% ethanol.

**Supplemental Figure S8.** Liquid chromatography-mass spectrometry analysis of a P-type *B. vulgaris* metabolite extract with 55% ethanol.

**Supplemental Figure S9.** Liquid chromatography-mass spectrometry analysis of an acidic hydrolyzed G-type *B. vulgaris* metabolite extract.

**Supplemental Figure S10.** Liquid chromatography-mass spectrometry analysis of an acidic hydrolyzed P-type *B. vulgaris* metabolite extract.

**Supplemental Figure S11.** Glucosylation activity of UGT73C9 to UGT73C13 toward G-type and P-type *B. vulgaris* sapogenin extracts.

**Supplemental Figure S12.** Overlaid Liquid chromatography-mass spectrometry analyses of metabolite extracts from the *B. vulgaris* plants used for the saponin abundance and UGT73C9 to -C13 expression correlation analysis.

**Supplemental Table S1.** Amino acid and nucleotide sequence identities of UGT73s used in the phylogenetic analysis.

**Supplemental Table S2.** Primers used in this study.

**Supplemental Data Set S1.** Structure elucidation of hederagenin and ole- anolic acid monoglucosides based on 1-D 1H- and 13C- and 2-D TOCSY-, COSY-, and HSQC-NMR data.

**Supplemental Data Set S2.** Multiple sequence alignment, amino acid se- quences, and nucleotide sequences used for the phylogenetic analysis.

**ACKNOWLEDGMENTS**

We are grateful to Rubini Kannangara for helpful discussions and commenting on the manuscript. Mohammed Siddik Motawie and Henrik Toft Simonsen are thanked for consulting in chemical aspects, and Tamara van Molkten is thanked for *P. nemorum* images. Peter McKenzie is acknowledged for naming the UGTs according to the UGT nomenclature. Mika Zagrobelny is thanked for help and discussions on the use of codeml in the PAML package. Received June 28, 2012; accepted September 30, 2012; published October 1, 2012.

**LITERATURE CITED**


