Crystal Structures of a Multifunctional Triterpene/Flavonoid Glycosyltransferase from Medicago truncatula

Hui Shao, Xianzhi He, Lahoucine Achnine, Jack W. Blount, Richard A. Dixon, and Xiaoqiang Wang

Plant Biology Division, Samuel Roberts Noble Foundation, Ardmore, Oklahoma 73401

Glycosylation is a ubiquitous reaction controlling the bioactivity and storage of plant natural products. Glycosylation of small molecules is catalyzed by a superfamily of glycosyltransferases (GTs) in most plant species studied to date. We present crystal structures of the UDP flavonoid/triterpene GT UGT71G1 from Medicago truncatula bound to UDP or UDP-glucose. The structures reveal the key residues involved in the recognition of donor substrate and, by comparison with other GT structures, suggest His-22 as the catalytic base and Asp-121 as a key residue that may assist deprotonation of the acceptor by forming an electron transfer chain with the catalytic base. Mutagenesis confirmed the roles of these key residues in donor substrate binding and enzyme activity. Our results provide an initial structural basis for understanding the complex substrate specificity and regiospecificity underlying the glycosylation of plant natural products and other small molecules. This information will direct future attempts to engineer bioactive compounds in crop plants to improve plant, animal, and human health and to facilitate the rational design of GTs to improve the storage and stability of novel engineered bioactive compounds.

INTRODUCTION

Glycosylation is quantitatively the most significant reaction on earth. It is catalyzed by a superfamily of enzymes, the glycosyltransferases (GTs), which have been classified into >70 families (Campbell et al., 1997) (http://afmb.cnrs-mrs.fr/CAZY/index.html). Family 1 GT enzymes, the UDP glycosyltransferases (UGTs), transfer UDP-activated sugar moieties to specific acceptor molecules. Twenty-seven UGTs have been identified in the human genome (de Wildt et al., 1999; Li et al., 2001); these play central roles in the metabolism and detoxification of foreign chemicals such as carcinogens and hydrophobic drugs. By contrast, plants contain a very large number of UGTs that are involved in the glycosylation of natural products (Vogt and Jones, 2000; Jones and Vogt, 2001) and contain a highly conserved consensus signature sequence called the PSPG motif (Hughes and Hughes, 1994; Paquette et al., 2003). Plants collectively synthesize >200,000 natural products, and a significant proportion of them are glycosylated. Consistent with this chemical complexity, 117 putative UGT genes have been identified in the genome of Arabidopsis thaliana (Ross et al., 2001). Likewise, >100 UGTs are predicted in the model legume Medicago truncatula based on an analysis of currently available EST and genomic sequences (Achnine et al., 2005).

Glycosylation is generally seen as a mechanism for the detoxification of xenobiotics or for facilitating the storage of bioactive molecules in plant vacuoles (Sandermann, 1992; Paquette et al., 2003). However, in some cases, glycosylated plant natural products exhibit higher bioactivity than their corresponding aglycones, as in the case of the triterpene saponins (Osborn, 2003). UGT71G1 was recently identified in M. truncatula as an enzyme functionally involved in the biosynthesis of saponins, although the enzyme exhibits more favorable kinetics with the flavonoid quercetin (Achnine et al., 2005). Triterpene glycosides have allelopathic, antimicrobial, insecticidal, and antiherbivory activities for plant protection (Osborn et al., 1996; Smail, 1996; Oleszek et al., 1999; Papadopoulou et al., 1999). They also have pharmacological, anticholesterolemic, anticancer, adjuvant, and hemolytic activities (Behboudi et al., 1999; Oh et al., 2000; Haridas et al., 2001), act on the cardiovascular, nervous, and digestive systems, and are used in sneezing powders, emetics, and cough syrups to facilitate expectoration (Dennehy, 2001). The saponins in M. truncatula are glycosides of the five triterpene aglycones soyasapogenol B, soyasapogenol E, medicagenic acid, hederagenin, and bayogenin (Huham and Sumner, 2002). These aglycones share similar structures, and UGT71G1 recognizes both medicagenic acid and hederagenin as acceptors (Figure 1) for the sugar moiety delivered from UDP-glucose. It is not yet clear which positions on the triterpene scaffold are glycosylated by UGT71G1; O-glycosylation at positions 3 and 28 is most common for this class of compound, but C-glycosylation of plant natural products has also been observed, and these products are of particular pharmaceutical interest (Clemens et al., 2004). Understanding the structural basis for the glycosylation of triterpene natural products will facilitate future metabolic engineering of these compounds to improve the health of plants (resistance to disease and herbivory), animals (via improved forage quality), and humans (through drug discovery).
To date, 17 members of the GT superfamily (but none from plants) have been analyzed structurally (Goldsmith et al., 1989; Vrielink et al., 1994; Charnock and Davies, 1999; Gastinel et al., 1999, 2001; Ha et al., 2000; Pedersen et al., 2000, 2003; Unigil et al., 2000; Persson et al., 2001; Gibbons et al., 2002; Gibson et al., 2002; Buschiazzo et al., 2004; Fritz et al., 2004; Lobosanov et al., 2004; Mulichak et al., 2004), and these structures fall into two different topologies/folds, GT-A and GT-B (Fritz et al., 2004; Bowles et al., 2005), which belong to the GT-B fold comprising two Rossmann-like domains. However, bacterial Gtfs share only ~10% sequence identity with plant UGTs, making it very difficult to model plant glycosylation processes without knowledge of plant UGT crystal structures.

Here, we report the structures of UGT71G1 from *M. truncatula* bound to UDP or UDP-glucose. These plant UGT crystal structures reveal the detailed interactions between the enzyme and its donor substrate as well as the roles of the UGT signature motif in substrate recognition and the catalytic activity of the enzyme. Further mutational and structural analyses, and comparisons with other GT structures, will provide a basis for understanding the complex and often overlapping substrate specificity and regiospecificity within the plant UGT superfamily.

### RESULTS

#### Overall Structure

The three-dimensional structure of UGT71G1 from *M. truncatula* bound with a UDP molecule was determined using the multi-wavelength anomalous dispersion (MAD) method with a crystal of Se-Met-substituted enzyme and refined to 2.0-Å resolution with a native data set. Data collection, phasing, and refinement statistics are presented in Table 1.

The structure of UGT71G1 consists of two N- and C-terminal domains with similar Rossmann-type folds and, as predicted, belongs to the GT-B fold (Figures 2 and 3). The N-terminal domain contains a central seven-stranded parallel β sheet flanked by eight α helices on both sides and a small two-stranded β sheet. The C-terminal domain contains a six-stranded β sheet flanked by eight α helices. The two domains pack very tightly and form a deep cleft with a UDP molecule bound.

The asymmetric unit contains two independent UGT71G1 molecules related by a noncrystallographic twofold axis. These two molecules are very similar to each other, with root mean square deviations (RMSD) of 1.04 and 1.52 Å for 456 Cα atoms and all atoms (residues 8 to 463), respectively. However, small differences were observed between the two molecules, especially in the region near the UDP binding site in the cleft between the N- and C-terminal domains, which included the C-terminal part of helix Nu1 (around residue Ile-20); loop Nβ2-Nu2 (around residue Pro-50), which is highly mobile with high B factors (~80 Å²) and poorly defined electron density in both molecules A and B; and the loop linker between the N- and C-terminal domains (around residue Asn-255), the electron density of which was poorly defined in molecule B. The helix Nu1 (P18 to H36) and loop Nβ2-Nu2 and the following helix Nu2 (K48 to Q66) in molecule B shift away slightly from the C-terminal domain, with the largest movement, 2.6 Å, for the Cα atom of Leu-23, which is adjacent to the key catalytic residue His-22 (see below). This suggests that the flexibility in the cleft region between the two subunits allows for movement to occur during substrate binding and catalysis.

#### Comparison with Other Members of the GT Superfamily

There are three known structures of family 1 GTs, all for bacterial vancomycin UDP glucosyltransferase Gtfs. Superimposing the structures of plant UGT71G1 and bacterial GtfD revealed strong structural similarity, with an RMSD of 2.23 Å for 242 Cα atoms, although the sequence identity is very low (~10%) (Figure 3). The largest difference between these two enzymes was observed in the region between the helix Nu5 and the strand Nβ6 (~155 to 210) in their N-terminal domains. A two-stranded β sheet is present in this region of the UGT71G1 structure, resembling an arm flipped to the opposite side near the helix Cα5 of the C-terminal domain, compared with the corresponding region in the GtfD structure. This region forms one side of the substrate (acceptor molecule) binding pocket, and this large difference in

---

**Figure 1.** Structures of Three Substrates of *M. truncatula* UGT71G1.

The known or predicted glycosylation sites are labeled with numbers: quercetin (3, 5, 7, 3', and 4'), medicagenic acid (3 or 28), and hederagenin (3 or 28).
Table 1. Data Collection, Phasing, and Refinement Statistics

<table>
<thead>
<tr>
<th>Data Collection and Phasing Statistics</th>
<th>Native (+ UDP)</th>
<th>Se peak (+ UDP)</th>
<th>Se inflection (+ UDP)</th>
<th>Se remote (+ UDP)</th>
<th>Complex (+ UPG)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution (Å)</td>
<td>2.0</td>
<td>2.4</td>
<td>2.4</td>
<td>2.4</td>
<td>2.6</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.5418</td>
<td>0.9797</td>
<td>0.98</td>
<td>0.9393</td>
<td>1.5418</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>62,825</td>
<td>36,121</td>
<td>35,553</td>
<td>34,461</td>
<td>29,545</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.2 (97.7)</td>
<td>94.4 (81.8)</td>
<td>97.0 (85.0)</td>
<td>92.4 (78.4)</td>
<td>99.8 (99.7)</td>
</tr>
<tr>
<td>Rsym (%)</td>
<td>4.1 (27.7)</td>
<td>8.2 (26.0)</td>
<td>5.7 (25.2)</td>
<td>6.0 (28.6)</td>
<td>5.9 (44.2)</td>
</tr>
<tr>
<td>L/σ (I)</td>
<td>13.7 (2.7)</td>
<td>16.5 (2.7)</td>
<td>16 (2.6)</td>
<td>16.1 (3.1)</td>
<td>12.1 (2.3)</td>
</tr>
<tr>
<td>Figure of merit</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enzyme-UDP</td>
<td>18.7</td>
<td>20.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enzyme-UDP-glucose (UPG)</td>
<td>22.9</td>
<td>28.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of protein atoms</td>
<td>7204</td>
<td>7204</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of solvent atoms</td>
<td>728</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of ligand atoms</td>
<td>50 (two UDP molecules)</td>
<td>72 (two UDP-glucose molecules)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average B factors (Å²)</td>
<td>31.8</td>
<td></td>
<td></td>
<td>56.3</td>
<td></td>
</tr>
<tr>
<td>RMSD from ideal values</td>
<td>0.0057</td>
<td>0.0070</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond length (Å)</td>
<td>1.27</td>
<td></td>
<td></td>
<td>1.29</td>
<td></td>
</tr>
</tbody>
</table>

Numbers in parentheses are for the highest-resolution shell.

*Rsym = \[\Sigma_{hkl} |I| - \langle I\rangle/\Sigma_{hkl}I\]

where I is the observed intensity and \(\langle I\rangle\) is the average intensity from observations of symmetry-related reflections. A subset of the data (8%) was excluded from the refinement and used to calculate the free R value (Rfree). R factor = \[\Sigma |F_{o}| - |F_{c}|/\Sigma |F_{o}|\].

conformation/folding changes the shape and size of the binding pocket. The comparison also shows that the similarity between the C-terminal domains of UGT71G1 and GtfD (RMSD = 2.13 Å for 137 Cα atoms) is higher than that between their N-terminal domains (RMSD = 2.34 Å for 120 Cα atoms), presumably because the C-terminal domains recognize similar donor molecules but the N-terminal domains recognize quite different acceptor molecules. The UGT71G1 structure is more compact in its N-terminal domain, presumably because its substrates, such as medicagenic acid and quercetin, are considerably smaller than the acceptor substrate desvancomycin vancomycin (DvV) of GtfD.

Comparison was also performed between the UGT71G1 structure and those of other GTs with the GT-B fold. *Escherichia coli* MurG (Coutinho et al., 2003) is closely related to UGT71G1 in the overall GT classification. Superimpositions between portions of the structures of UGT71G1 and MurG gave RMSD values of 2.27 Å (199 Cα atoms of the overall structures), 2.23 Å (95 Cα atoms in the N-terminal domains), and 1.95 Å (110 Cα atoms in the C-terminal domains).

Trehalose-6-phosphate synthetase OtsA, T4 bacteriophage β-glucosyltransferase (BGT), and glycogen synthase are relatively distant from UGT71G1 in classification, although they use the same UDP-sugar or, in the case of glycogen synthase, the related ADP-sugar. Superimposing the overall structures of UGT71G1 and OtsA, which have very little sequence identity, gave an RMSD of 2.21 Å (158 Cα atoms). Superimposing their N- and C-terminal domains gave RMSD values of 2.17 Å (96 Cα atoms) and 2.27 Å (70 Cα atoms), respectively; the N-terminal rather than the C-terminal domain is more similar to the corresponding domain of UGT71G1, in contrast with the situation with UGT71G1 and GtfD or MurG. Similarly, superimposing the N- and C-terminal domains of UGT71G1 and BGT gave RMSD values of 2.28 Å (79 Cα atoms) and 2.29 Å (70 Cα atoms), respectively.

Interactions between Enzyme and Donor Ligands

Although UDP-galactose (a nondonor) and quercetin were included for cocryostallization with UGT71G1, the only substrate electron density observed was for the UDP moiety of UDP-galactose (Figure 4A) as a result of possible hydrolysis. This is similar to the situation with the bacterial GtfD structure, in which only GDP was observed, although GtfD was cocryostallized with GDP-glucose (Mulichak et al., 2004). Soaking crystals of UGT71G1 with UDP-glucose resulted in a weak electron density map for the glucose portion of the molecule in the enzyme structure at 2.6 Å (Figure 4B). The B factor of the glucose moiety is higher than that of the UDP moiety, indicating either flexibility or low occupancy of the sugar attributable to the possible hydrolysis of UDP-glucose.

UDP or UDP-glucose is almost completely buried in a long, narrow channel mainly within the C-terminal domain of the enzyme. In the 2.0-Å structure of UGT71G1 complexed with UDP, the enzyme interacts with UDP mainly through residues in the UGT signature PSPG motif (residues 339 to 382) (Figure 4C, Table 2). The uracil ring of UDP interacts through hydrogen bonds between the N3 and O4 atoms of the ring and the main chain oxygen and nitrogen atoms of Ala-340 and also forms parallel π-stacking interactions with the indole ring of Trp-339.
The ribose ring of UDP makes contacts with the enzyme through hydrogen bonds between its O2* and O3* atoms and the OE1 and OE2 atoms of Glu-365 and between its O2* atom and the NE2 atom of Gln-342. The α-phosphate forms hydrogen bonds through its O1A atom and the ND2 atom of Asn-361, its O2A atom and the OG and N atoms of Ser-362, and its O3A atom and the NE2 atom of His-357. The β-phosphate forms hydrogen bonds through its O3B oxygen and the OH atom of Tyr-379 and the NE2 atom of His-357 and between its O2B oxygen and the OG and N atoms of Ser-285, the only interacting residue outside of the UGT signature motif region.

Interactions between the enzyme and the glucose moiety of the donor are observed in the 2.6-Å structure of UGT71G1 in its complex with UDP-glucose (Figure 4C, Table 2). The O2* and O3* atoms of the glucose ring form hydrogen bonds with the NE2 atom of Gln-382 and the OE1 atom of Glu-381. The O4* atom of the glucose ring contacts both the OE2 atom of Glu-381 and the N atom of Trp-360. On the other hand, several interactions observed between UDP and the enzyme are not present in the model of UDP-glucose bound to the enzyme. A small rotation and shift in the position of the β-phosphate of UDP-glucose weakens the interaction of the O2B and O3B atoms of β-phosphate and the O3A atom of α-phosphate with protein residues.

In UGT71G1 binding UDP-glucose, GtfD binding TDP, trehalose-6-phosphate synthetase OtsA binding UDP, MurG binding UDP-GlcNAc, and T4 bacteriophage BGT and its mutant D100A binding UDP or UDP-glucose, the sugar donor molecules are observed in a similar region of the enzymes (Gibson et al., 2002; Hu et al., 2003; Lariviere et al., 2003; Mulichak et al., 2004). The PSPG motif of UGT71G1 and the corresponding sugar binding regions in these different GTs share a very similar conformation, αβ/α/β, even though the sequence identities are very low and the detailed interactions with the sugar donors also vary. Superimpositions show that the UGT signature motif of UGT71G1 and the corresponding regions of GtfD and MurG share higher structural similarity, indicative of similar binding modes with sugar donors, and differ from those for OtsA and BGT, in which the first β strands of their αβ/αβ/β motifs are long and the second β strands cannot interact with sugar donors (Figure 5).

The Acceptor Binding Site

A very narrow, deep pocket is observed adjacent to the UDP binding site, mainly consisting of residues from the N-terminal domain (Figure 6), which would appear to be the binding pocket for acceptor molecules. The location of this pocket is similar to that of the substrate binding site observed in the structure of GtfD bound with DVV and TDP (Mulichak et al., 2004). In the GtfD structure, Asp-13 is close to both DVV and TDP and has been proposed as the general base for the catalysis of glycosyl transfer. His-22 occupies the corresponding position in UGT71G1 and is 4.7 Å from the C1′ atom of UDP-glucose, as shown in the structure of UGT71G1 (molecule A) complexed with UDP-glucose. Close to it (~3 Å), there is an acidic residue (Asp-121) that may form an electron transfer chain to His-22 to help deprotonate acceptor molecules. The sequence alignment of multiple plant UGTs (Figure 7) shows that corresponding His and Asp residues are conserved in most plant UGTs.

There are several potential O-glycosylation sites (five hydroxyl groups) on the quercetin molecule, and multiple products are formed after incubation of quercetin and UDP-glucose with UGT71G1 (Figure 8A). All five products from the in vitro incubation of recombinant UGT71G1 with UDP-glucose and quercetin exhibited a similar UV spectrum to that of quercetin (Figure 8B), had identical molecular ion masses (462.5 to 462.9 D) consistent with quercetin monoglucoside structures, and disappeared after incubation with β-glucosidase (data not shown). This indicates that, in spite of the production of one favored product (peak 5 in Figure 8A, identified as quercetin-3′-O-glucoside), UGT71G1 is capable of glucose transfer to each of the five hydroxyl groups on the quercetin molecule. To probe the structural basis of the apparently promiscuous acceptor binding to UGT71G1, molecular docking was performed using the...
Figure 3. Comparison of *M. truncatula* UGT71G1 and *A. orientalis* GtfD.

(A) Structure-based sequence alignment of *M. truncatula* UGT71G1 and *A. orientalis* GtfD. The secondary structure elements observed in the UGT71G1 structure are shown above the alignment. The UGT signature motifs are enclosed in a green box. Conserved residues are highlighted. This figure was produced with ENDscript (Gouet and Courcelle, 2002).

(B) Stereo diagram showing the superimposition of the structures of UGT71G1 (molecule A; brown) and GtfD (blue; Protein Data Bank [PDB] code 1RRV).
The docking in the active site of the hydroxyl groups at positions 3 or 3’ of quercetin gives fitness scores slightly higher than for the other three hydroxyls, and also more potential interactions between quercetin and the protein and donor. In the model with the 3’-OH docked in the active site (Figure 6A), hydrogen bonds may form between the 3-OH of quercetin and the O1B atom of UDP-glucose and between the 5-OH of quercetin and the OG atom of Ser-285.

The triterpene substrates hederagenin and medicagenic acid are more hydrophobic and significantly larger than quercetin. Compared with the docking results with quercetin, docking with medicagenic acid and hederagenin gave lower fitness scores and weaker interaction with the enzyme, consistent with previously reported kinetic data (Achnine et al., 2005). The molecular docking study for hederagenin and medicagenic acid also suggested that O-glycosylation at position 3 would be preferred over other positions. In the model docked with hederagenin (Figure 6B) or medicagenic acid, the OH group at position 3 points to the NE2 atom of His-22 (~3.2 Å) and is located just above the C1’ atom of UDP-glucose (~3.8 Å). Similar to the proposed general mechanism of other inverting GTs (Unligil and Rini, 2000), His-22 would act as an effective general base to deprotonate the OH group of the acceptor molecule, then the newly formed nucleophilic oxygen would attack the C1’ carbon of the UDP-glucose directly to displace the UDP moiety. The carboxyl groups at position 28 of the triterpene acceptors may hydrogen bond with the main chain oxygen atom of Gly-51. The residues Pro-18, Phe-49, Pro-53, Phe-54, Leu-85, Pro-88, Ile-92, and Phe-123 provide a hydrophobic environment for the acceptor substrates. A solution of the model for docking the carboxyl oxygen atom at position 28 of hederagenin to the active site had a slightly lower fitness score than that for the 3-hydroxyl. In the former case, the hydroxyl group at position 3 may interact with the main chain oxygen atom of Gly-51.

**Figure 4. Donor Molecules and Their Interactions with UGT71G1.**

(A) A |F<sub>obs</sub>| − |F<sub>calc</sub>| electron density omit map of bound UDP contoured at 1.5 σ is superimposed on a ball-and-stick model of the UDP molecule. (B) A |F<sub>obs</sub>| − |F<sub>calc</sub>| electron density omit map of bound UDP-glucose contoured at 1.5 σ is superimposed on a ball-and-stick model of the UDP-glucose molecule. (C) Stereo diagram showing interactions between bound UDP-glucose and UGT71G1 side chains. The structure of UDP-glucose is shown as a ball-and-stick model. Hydrogen bonding interactions are indicated by dashed lines. Interactions observed only between UDP-(galactose) and UGT71G1 in the 2.0-Å structure are indicated by green dashed lines.

**Mutagenesis of UGT71G1**

To further explore the roles of the potential key residues identified above, structure-directed point mutants of UGT71G1 were constructed, targeting residues involved in direct contact with the sugar donor, the proposed general base His-22, and its partner, Asp-121. Substitution of His-22 with Ala (H22A) led to the complete loss of enzyme activity, which is consistent with the proposed role of His-22 as the general base. Substitution of Asp-121 with Ala or Asn resulted in no detectable enzyme activity. This is similar to the situation with trypsin, for which the mutant D102N (in the catalytic triad) showed a large loss in activity (Sprang et al., 1987). Also similar to the case of trypsin, Asp-121 in UGT71G1 is nearly buried and forms a hydrogen bond with a catalytic His residue. Our results demonstrate that Asp-121 and its proper interaction with His-22 are essential for enzyme activity. In the model docked with acceptors, Asp-121 is not able to contact the acceptor molecules directly. This finding suggests that Asp-121 may play an important role in assisting His-22 by initialization or activation of the reaction through electron transfer. The E381A mutant also showed no detectable activity, demonstrating the key role of Glu-381 in recognition of the sugar donor substrate because the side chain of Glu-381 forms hydrogen bonds with the O3’ and O4’ atoms of the glucose ring of the donor (Figure 4C). By contrast, mutant A340F retained...
activity similar to that of the wild-type enzyme reported previously (Achnine et al., 2005). Ala-340 forms hydrogen bond interactions with the uracil ring of the donor through its main chain oxygen and nitrogen atoms; therefore, such bonds will still be supplied by Phe.

**DISCUSSION**

**Comparisons between GT-B and GT-A Fold GTs**

Comparison of the three-dimensional structures of UGT71G1 and several other GT-B fold enzymes revealed high similarity despite the very low sequence identity. Gtfs and MurG, with a similar donor binding motif, share higher similarity in their C-terminal domains than in their N-terminal domains, suggesting a similar sugar donor binding mode and probably a common evolutionary origin. OtsA and BGT, with a slightly different donor binding motif, share higher similarity in their N-terminal domains with UGT71G1. Glycogen synthase (PDB code 1RZV) shares higher similarity with UGT71G1 in the C-terminal domain, but its donor binding region is closer to that of OtsA or BGT (data not shown). This suggests that these enzymes have different sugar donor binding modes and evolutionary origins.

GT-A fold enzymes contain only a single extended domain (or two tightly associated and abutting domains) that recognizes both donor and acceptor (Davies et al., 2001; Coutinho et al.,

---

**Table 2. Interactions between Enzyme Molecules (A or B) and Donors (UDP or UPG)**

<table>
<thead>
<tr>
<th>Protein Residue: Atom</th>
<th>Donor Atom</th>
<th>Distance (Å)</th>
<th>Molecule A, UDP</th>
<th>Molecule B, UDP</th>
<th>Molecule A, UPG</th>
<th>Molecule B, UPG</th>
</tr>
</thead>
<tbody>
<tr>
<td>A340: N</td>
<td>O4</td>
<td>3.0</td>
<td>3.1</td>
<td>2.8</td>
<td>3.3*</td>
<td></td>
</tr>
<tr>
<td>O</td>
<td>O3, O2</td>
<td>2.7, 3.3</td>
<td>2.6, 3.3</td>
<td>2.6, 3.1</td>
<td>2.6, 2.9</td>
<td></td>
</tr>
<tr>
<td>Q342: NE2</td>
<td>O2*</td>
<td>3.0</td>
<td>3.0</td>
<td>3.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E365: OE1</td>
<td>O2*</td>
<td>2.7</td>
<td>2.6</td>
<td>2.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS2</td>
<td>O2*, O3*</td>
<td>3.2, 2.8</td>
<td>3.2, 2.8</td>
<td>3.1, 2.9</td>
<td>3.0, 2.6</td>
<td></td>
</tr>
<tr>
<td>N361: ND2</td>
<td>O1A</td>
<td>3.1</td>
<td>3.2</td>
<td>3.1</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>S362: N</td>
<td>O2A</td>
<td>2.8</td>
<td>2.9</td>
<td>3.1</td>
<td>2.9</td>
<td></td>
</tr>
<tr>
<td>OG</td>
<td>O2A</td>
<td>2.5</td>
<td>2.8</td>
<td>2.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>H357: NE2</td>
<td>O3A</td>
<td>2.9</td>
<td>2.9</td>
<td>3.4</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>OG</td>
<td>O3B</td>
<td>3.0</td>
<td>3.2</td>
<td>3.4</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>H357: NE2</td>
<td>O3B</td>
<td>3.0</td>
<td>3.2</td>
<td>3.4</td>
<td>3.4</td>
<td></td>
</tr>
<tr>
<td>Y379: OH</td>
<td>O3B</td>
<td>2.4</td>
<td>2.6</td>
<td>3.3 (w. O2B)</td>
<td>3.2 (w. O2B)</td>
<td></td>
</tr>
<tr>
<td>S285: N</td>
<td>O2B</td>
<td>2.7</td>
<td>2.9</td>
<td>3.4</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>OG</td>
<td>O2B</td>
<td>2.7</td>
<td>2.7</td>
<td>3.5</td>
<td>3.3</td>
<td></td>
</tr>
<tr>
<td>W360: N</td>
<td>O4’</td>
<td>2.9</td>
<td>2.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E381: OE1</td>
<td>O3’</td>
<td>2.4</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OE2</td>
<td>O4’</td>
<td>2.6</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q382: NE2</td>
<td>O2’</td>
<td>2.7</td>
<td>2.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Numbers for distances >3.3 Å are underlined.

---

**Figure 5. Comparison of Donor Binding Regions of UGT71G1 and Other GT-B Fold Enzymes.**

Stereo diagram showing superimposition of the UGT signature motif regions of UGT71G1 (yellow), GtfD (red), MurG (green; PDB code 1F0K), OtsA (cyan; PDB code 1GZ5), and BGT (black; PDB code 1J39).
2003). The structures of GT-A fold enzymes and the N- and C-terminal domains of GT-B fold enzymes are similar to each other. Superimposition between a GT-A enzyme, human β1,3-glucuronidtransferase I (family 43; PDB code 1FGG) (Pedersen et al., 2000), and each domain of UGT71G1 gave RMSD values of 2.3 Å for 80 Ca atoms in the N-terminal domain and 1.96 Å for 40 Ca atoms in the C-terminal domain. This finding suggests that GT-A and GT-B enzymes are evolutionarily related. During evolution, two domain enzymes were probably produced, with one domain recognizing similar donors and another adapting to divergent acceptors in the natural environment.

**Substrate Binding and Catalysis by UGT71G1**

Kubo et al. (2004) reported that a single mutation, H374Q, in *Aralia cordata* UDP-galactose:anthocyanin galactosyltransferase (corresponding to Gln-382 in UGT71G1) switched the preferred donor from UDP-galactose to UDP-glucose and suggested that the last residue, Gln or His, in this signature motif would define the sugar donor specificity. However, the *Scutellaria baicalensis* UDP-glucose:flavonoid glycosyltransferase mutant Q382H does not acquire galactosyltransferase activity, suggesting that other residues might also play roles in sugar recognition and specificity in plant UGTs. In the structure of UGT71G1 bound with UDP-glucose, the O2' and O3' atoms of the glucose ring of the donor form hydrogen bonds with the NE2 atom of Gln-382 and the OE1 atom of Glu-381, respectively. The O4' atom of the glucose ring contacts both the OE2 atom of Glu-381 and the N atom of Trp-360. If a UDP-galactose is placed in the position of the UDP-glucose in the structure of UGT71G1, the O4' atom of galactose points away from the OE2 atom of Glu-381 and can no longer form a hydrogen bond. This suggests that Glu-381 may be a key residue for sugar recognition and specificity.

A structural model for catalysis by plant UGTs has been proposed based on a modeling and mutagenesis study of UGT73A5, a betanidin glucosyltransferase from *Dorotheanthus bellidiformis* (Hans et al., 2004). However, the low identity (~14%) between UGT73A5 and the bacterial GtfB used as the basis for the homology modeling made it difficult to model regions with insertion or deletion of amino acid residues. Nevertheless, His-22 was identified as a key residue for enzyme activity, although it was oriented differently from that shown in the crystal structure of UGT71G1, such that the acceptor binding pocket was predicted to be in a different location from that suggested by our docking results. The position of the binding pocket for the sugar donor was predicted correctly, but the sugar

---

**Figure 6. The Putative Acceptor Binding Pocket.**

Stereo diagram showing quercetin (A) and hederagenin (B) docked into the proposed binding pockets. Quercetin, hederagenin, and UDP-glucose are shown as ball-and-stick models. Some protein residues in the acceptor binding pocket are labeled and shown in cyan as bond models. Distances (Å) between the OH group of acceptors and the atom NE2 of His-22 or the atom C1' of UDP-glucose are labeled and indicated with dashed lines.
Figure 7. Comparison of the Donor and Acceptor Regions of Functionally Characterized Plant UGTs.

Sequence alignment of 39 plant UGTs, showing the acceptor binding region (first three fragments: 1 to 55, 81 to 95, and 117 to 151) and the donor binding region (last two fragments: 282 to 286 and 339 to 382). Identical residues are highlighted, and similar residues are enclosed in boxes. Residues His-22 and Asp-121 are marked with asterisks. The alignment was performed using ClustalX (Thompson et al., 1997). The plant UGTs include Medicago truncatula UGT71G1 and UGT73K1; Allium cepa UGT73G1 and UGT73J1; Aralia cordata GaT; Bellis perennis UGT1; Brassica napus SGT1; Catharanthus roseus UGT2; Citrus maxima 1,2RhaT; Dorotheanthus bellidiformis UGT73A5; Gentiana triflora GT; Glycyrrhiza echinata UGT73F1; Nicotiana tabacum GT1a, GT2, and GT3; Phaseolus lunatus ZOG1; Phaseolus vulgaris ZOX1; Sorghum bicolor UGT85B1; Stevia rebaudiana UGT74G1, UGT76G1, and UGT85C2; and Zea mays cisZOC1 and cisZOG2. All others are from Arabidopsis thaliana.
was predicted to lie in the opposite orientation to that revealed from our crystal structure, and His-370 (His-357 in UGT71G1) interacted with the phosphates rather than the uracil ring of the donor molecule. Based on the modeling study, Glu-394 was predicted to be part of a catalytic triad. However, as revealed in the crystal structure of UGT71G1, Glu-381 (corresponding to Glu-394 in UGT73A5) contacts the glucose moiety of the donor and plays an important role in sugar recognition and specificity. Rather than Glu-381, the highly conserved Asp residue (Asp-121) located in the acceptor binding pocket of UGT71G1 may be essential for enzyme catalysis.

In the model revealed by our crystal structures, Asp-121 interacts with His-22 at a distance of ~3 Å. His-22 is close to the glucose ring of the sugar donor and the docked acceptor. The imidazole group of His-22 acts as a general base to deprotonate an OH group on the acceptor molecule and transfers the proton to Asp-121. The deprotonated acceptor, as a nucleophilic oxyanion, can then attack at the C1' carbon center of the UDP-glucose from the opposite side of the sugar ring, with displacement of the UDP moiety. The crystal structure of UGT71G1 presented here thus provides a basis for understanding the catalytic mechanism for glycosyl transfer and also provides a starting point for homology modeling of diverse plant UGTs.

**Substrate Specificity in the Plant UGT Superfamily**

Most of the residues in direct contact with donor molecules, such as Trp-339, Gln-342, His-357, and Glu-365, are conserved in all 39 plant UGTs functionally characterized to date (Figure 7). In the position corresponding to Glu-381 of UGT71G1, Asp is frequently observed, which may enter into a similar interaction with the donor. At the position corresponding to Gln-382 of UGT71G1, Gln is conserved in glucosyltransferases, His is present and conserved in galactosyltransferases, and an Asn residue is observed in Arabidopsis UGT78D1, which uses UDP-rhamnose as the sugar donor. Residues at these two positions interact with the sugar moiety of the donor and therefore would be predicted to be critical for sugar specificity.

The apparent lack of stringent structural requirements for acceptor binding and catalysis is consistent with the evolution of varied and broad substrate specificities of many plant GTs (Vogt and Jones, 2000; Bowles et al., 2005). Of five UGTs that we have characterized recently from *M. truncatula* (Achnine et al., 2005), all have multiple in vitro substrates, suggesting multiple potential in vivo substrates. UGT71G1 can catalyze the glucosylation of two triterpenes (medicagenic acid and hederagenin), the flavonol quercetin, and the isoflavones genistein and biochanin A. For assessment of in vivo function, therefore, it is necessary to understand which potential substrates are present in the cell types in which the enzyme is expressed. In the case of methyl jasmonate–induced *Medicago* cell suspension cultures, such analysis points to UGT71G1 glycosylating medicagenic acid rather than the catalytically preferred substrates biochanin A or quercetin (Achnine et al., 2005).

When considered in terms of the multiple products formed from each potential substrate, such apparent promiscuity raises two important questions: why do higher plants possess so many different UGTs, and how are their activities regulated in vivo to

---

**Figure 8.** Glucosylation of Quercetin by Recombinant UGT71G1.  
(A) HPLC trace showing the substrate quercetin (Q) and the five products (numbered 1 to 5). Solid line, complete reaction mixture; dashed line, standards for quercetin (Q), quercetin-3-O-glucoside (peak 3), and quercetin-4'-O-glucoside (peak 4). mAU, milli-absorption unit.  
(B) UV absorption spectra of quercetin (Q) and products 1 to 5.
allow for the production of a more limited number of products than suggested by their in vitro substrate/product specificities. The importance of UGTs for the detoxification of xenobiotics (allelochemicals before the human introduction of synthetic herbicides and pesticides) may have driven the evolution of new forms of the enzyme differing in residues around the acceptor binding pocket.

It has been suggested that UGTs may be associated in enzyme complexes with the early enzymes of the pathway in which they are involved, and such metabolic channeling would essentially limit a particular GT to a specific pathway irrespective of the enzyme’s in vitro substrate specificity (Jorgensen et al., 2005). The presence and/or nature of physical interactions between any UGT and a related biosynthetic enzyme remain unknown. The structures described in this work provide a basis for the exploration of such interactions.

METHODS

Expression and Purification of UGT71G1

The UGT71G1 cDNA from Medicago truncatula (Achnine et al., 2005) was cloned into the pET28a expression vector (Novagen) with a hexahistidine tag and a thrombin cleavage site. Escherichia coli BL21(DE3) cells transformed with the plasmid were grown at 37°C in Luria-Bertani medium containing 50 μg/mL kanamycin until AoD = 0.6 to 0.8. Cultures were induced with 0.5 mM isopropyl-1-thio-β-D-galactopyranoside and grown overnight at 16°C. Cells were pelleted and resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10 mM imidazole, and 10 mM β-mercaptoethanol). After lysis with a French press and centrifugation at 12,000 rpm at 4°C for 20 min, Ni²⁺-nitritriacetic acid agarose was added to the supernatant containing the target proteins. After incubation for 40 to 60 min, the mixture was transferred into a disposable column and washed extensively with lysis buffer (~50 column volumes). The His-tagged proteins were eluted with elution buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 250 mM imidazole, and 10 mM β-mercaptoethanol). Incubation with biotinylated thrombin during overnight dialysis at 4°C against 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, and 10 mM β-mercaptoethanol cleaved the N-terminal His tag. Diazymed proteins were incubated with streptavidin agarose and Ni²⁺-nitritriacetic acid agarose to remove thrombin and the cleaved His tag, respectively. The proteins were further purified on Resource Q and Superdex-200 gel filtration columns (Amersham Pharmacia Biotech) and concentrated to 5 mg/ml in 10 mM NaCl, 10 mM Tris-HCl, pH 7.5, and 5 mM β-mercaptoethanol.

Se-Met-substituted UGT71G1 was purified using the recombinant protein in E. coli B834(DE3) cells (Novagen) grown in M9 minimal medium supplemented with Se-Met (Doubié, 1997). The procedure for expression and purification was similar to the procedure described above for the native protein.

Crystallization and Data Collection

Crystals of UGT71G1 were grown from hanging drops by the vapor diffusion method. UGT71G1 protein at a concentration of 5 mg/ml was mixed with 5 mM UDP-galactose and 5 mM quercetin at a 2:1 (v/v) ratio, then mixed with an equal volume of reservoir solution containing 40% (v/v) polyethylene glycol 3350, 0.2 M ammonium acetate, and 0.1 M sodium citrate, pH 5.6. The mixture was equilibrated over the reservoir solution at 20°C. Crystals typically grew over 2 to 5 d with dimensions of 0.1 × 0.1 × 0.2 mm.

Before data collection, the crystals were flash-cooled to ~180°C. Data from a native protein crystal were measured to 2.0 Å with a Raxis IV++ image plate detector and a RU3H x-ray source. The crystal belonged to space group P2₁ (a = 53.7 Å, b = 90.6 Å, c = 101.8 Å, β = 102.6°). There were two molecules per crystallographic asymmetric unit, with 47% solvent content and a Matthews coefficient Vm of 2.6 Å³/D. A 2.4-Å MAD data set was collected from a Se-Met derivative crystal, using three wavelengths, was collected at the Gulf Coast Protein Crystallography Consortium beamline using a MAR CCD detector.

Cryostals of UGT71G1 complexed with UDP-glucose were obtained by soaking the crystals described above with mother liquid containing saturated UDP-glucose. Five changes of this solution were made during the 3-d soaking period, the last being a few hours before freezing the crystal for data collection. A 2.6-Å diffraction data set was collected with a Raxis IV++ image plate detector and a RU3H x-ray source.

All data sets were processed using denzo and scalepack or HKL2000 (Otwinowski and Minor, 1997).

Structure Determination and Refinement

The structure of UGT71G1 was determined using the MAD method. The MAD data (40 to 2.4 Å) were analyzed with SOLVE (Terwilliger and Berendzen, 1999), and 20 of 24 Se sites were located, yielding an overall figure of merit of 0.44. The program RESOLVE (Terwilliger, 2000) was used for electron density modification, phase extension to 2.0 Å by combining with the native data, and automated model building. Interactive model building and crystallographic refinement were performed using the programs O (Jones et al., 1991) and CNS (Brünger et al., 1998), respectively. A bulk solvent correction was applied. B factors were refined individually for the structure of UGT71G1 complexed with UDP at 2.0 Å, and grouped B factor refinement was performed for the structure of UGT71G1 complexed with UDP-glucose at 2.6 Å. Water molecules were added with Arp/wArp (Lamzin et al., 2001) and checked manually for inclusion. In the models, the first two and seven amino acid residues in the N terminus were not observed in molecules A and B, respectively, and the last two residues in the C terminus were disordered in both molecules.

The program PROCHECK (Laskowski et al., 1993) was used to check the models. All backbone ψ-ω torsion angles are within allowed regions of the Ramachandran plot.

Molecular Docking

The automated docking program GOLD was used to dock acceptors into the UGT71G1 active site (Jones et al., 1997). Default genetic algorithm parameters for controlling the operation of the docking process were used. All docking calculations were restricted to the predicted binding pocket by defining the active site with residue His-22. For quercetin, medicagenic acid, and hederaogenin, there are several potential glycosylation sites, and a distance constraint was used to define each potential site for calculation. GOLDscore was used to identify the lowest energy docking results. The hydrogen bonds and van der Waals contacts between ligands and enzyme were analyzed to identify the optimal binding mode.

Mutagenesis and Enzyme Assay

Site-directed mutants of UGT71G1 were constructed using the Quik-Change strategy (Stratagene). The mutant proteins were expressed and purified using the same procedures described above for native protein except for the thrombin cleavage. Enzyme assay was performed essentially according to a reported method (Lim et al., 2003). Each GT activity assay mix (200 μL) contains 2 μg of recombinant protein, 50 mM Tris-HCl, pH 7.0, 14 mM β-mercaptoethanol, 5 mM UDP-glucose, and 1 mM quercetin. For mutants, increasing concentrations of substrates (2 mM quercetin and 10 mM UDP-glucose, or 5 mM quercetin and 25 mM UDP-glucose) and proteins (4 μg) were also used. The reaction was performed.
at 30°C for 1 h and stopped by the addition of 20 μL of trichloroacetic acid, quick-frozen, and stored at –20°C before HPLC analysis with a Waters Spherisorb 5u ODS2 C18 reverse-phase column (250 × 4.6 mm) on an Agilent HP1100 HPLC device equipped with an autosampler, a quaternary pump, and a diode array detector. Elution was as described previously (Achnine et al., 2005).

**Accession Numbers**

The coordinates and structure factors for the structures of UGT71G1 in complexes with UDP and UDP-glucose have been deposited to the PDB with codes 2ACV and 2ACW, respectively. The plant UGTs included in the alignment (Figure 7) and their accession numbers are as follows: *Medicago truncatula* UGT71G1 (accession number AAW56092) and UGT73K1 (AAW56091); *Allium cepa* UGT73G1 (AAP88406) and UGT73J1 (AAP88407); *Arabidopsis thaliana* UGT73K1 (AAW56091); *Bellis perennis* UGAT (BAD77944); *Brassica napus* SGT1 (AAF88390); *Catharanthus roseus* UGT2 (BAD29722); *Citrus maxima* 1,2RhaT (AAAL06646); *Crocus sativus* UGTCs2 (AAP94878); *Dorotheanthus bellidiformis* UGT73A5 (CAJ36231); *Gentiana triflora* 3’GT (BAC54092); *Glycyrrhiza echinata* UGTT3F1 (BAC78438); *Nicotiana tabacum* GT1a (BAB60721), GT2 (BAB88935), and GT3 (BAB88934); *Phaseolus vulgaris* UGT74G1 (AAR06920), in native and nucleotide-complexed forms. *Botanicals in cardiovascular health*. Clin. Obstet. Gynecol. 44, 814–823.

**REFERENCES**


Crystal Structures of a Multifunctional Triterpene/Flavonoid Glycosyltransferase from Medicago truncatula

Hui Shao, Xianzhi He, Lahoucine Achnine, Jack W. Blount, Richard A. Dixon and Xiaoqiang Wang

Plant Cell 2005;17;3141-3154; originally published online October 7, 2005;
DOI 10.1105/tpc.105.035055

This information is current as of June 20, 2017