Mutational Analysis of the Medicago Glycosyltransferase UGT71G1 Reveals Residues That Control Regioselectivity for (Iso)flavonoid Glycosylation*

Received for publication, June 15, 2006, and in revised form, August 31, 2006. Published, JBC Papers in Press, September 18, 2006, DOI 10.1074/jbc.M605767200

Xian-Zhi He, Xiaqiang Wang, and Richard A. Dixon

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

The plant glycosyltransferase UGT71G1 from the model legume barrel medic (Medicago truncatula) glycosylates flavonoids, isoflavonoids, and triterpenes. It can transfer galactose to each of the five hydroxyl groups of the flavonol quercetin, with the 3′-O-glucoside as the major product, and to the A-ring 7-hydroxyl of the isoflavone genistein. The sugar donor and acceptor binding pockets are located in the N and C termini, respectively, of the recently determined crystal structure of UGT71G1. The residues forming the binding pockets of UGT71G1 were systematically altered by site-directed mutagenesis. Mutation of Phe148 to Val, or Tyr202 to Ala, drastically changed the regioselectivity for quercetin glycosylation from predominantly the 3′-O-position of the B-ring to the 3-O-position of the C ring. The Y202A mutant exhibited comparable catalytic efficiency with quercetin to the wild-type enzyme, whereas efficiency was reduced 3–4-fold in the F148V mutant. The Y202A mutant gained the ability to glycosylate the 5-hydroxyl of genistein. Additional mutations affected the relative specificities for the sugar donors UDP-galactose and UDP-glucuronic acid, although UDP-glucose was always preferred. The results are discussed in relation to the design of novel biocatalysts for production of therapeutic flavonoids.

Glycosylation is among the most important chemical reactions in plants. It involves transfer of a nucleotide diphosphate-activated sugar moiety to an acceptor molecule and is catalyzed by a diverse group of uridine diphosphate glycosyltransferases (UGTs). All major classes of plant secondary metabolites, including flavonoids (1, 2), terpenoids (3–5), and alkaloids (6), can be modified by glycosylation, and more than 300 different flavonoid glycosides have been identified from plants (7).

Glycosylation of plant natural products has large effects on their solubility, chemical properties, compartmentation, storage, and biological activity (8, 9). In Gentiana triflora, the blue flower pigmentation results from a delphinidin-3,5,3′-tri-O-glucoside (10). Skin color in grape varieties (11), bitterness in citrus (12, 13), and sweetness in Stevia leaves (14) are due to the glycosylation of anthocyanin, flavonol (narigenin), or terpene (stevioside), respectively. Furthermore, glycosylation of secondary metabolites impacts disease resistance and plant development (15, 16). Considering the broad spectrum of secondary metabolite acceptors, and the many different types of nucleotide diphosphate sugar donors (13, 14, 17–19), UGTs can be seen as central players in determining chemical diversity in plants.

The flavonol quercetin is a common constituent of plants. It has five hydroxyl groups, at positions of 3, 5, 7, 3′, and 4′ (Fig. 1A). As one of the major antioxidants derived from plant sources, quercetin is beneficial to human health and has been ascribed anticancer (20–22), anti-inflammatory (23, 24), and anti-allergic activity (25, 26). In nature quercetin is often glycosylated on one or more of the five hydroxyl groups to increase its solubility, stability, and bioavailability. The position of conjugation has a significant impact on its biological activity and potential health benefits for humans (20, 21).

Recent studies have suggested that many plant secondary metabolite UGTs are more regiospecific than substrate specific (11, 22–26). Recently, UGTs from Arabidopsis thaliana were extensively analyzed both in vitro and in vivo with quercetin as acceptor; from a total of 29 UGTs that exhibited activity with quercetin, 14 catalyzed glycosylation either on the C-ring 3′-hydroxyl or the A-ring 7-hydroxyl, the most commonly reported glycosylation positions for quercetin derivatives isolated from plants. Eleven of the UGTs produced two products, glycosylating the 7- and 3′-hydroxyls, 7 and 4′-hydroxyls, or 3- and 4′-hydroxyls. The remaining enzymes produced three or four monoglucosides substituted on different hydroxyl groups (27). Some of these UGTs also glycosylated additional substrates such as caffeic acid (UGT71C1) (28), indole-3-acetic acid (UGT84B1) (29), or salicylic acid (UGT74F1) (24).

A single amino acid mutation in a UDP-galactose:anthocyanin galactosyltransferase from Aralia cordata was reported to alter sugar donor specificity (18). However, the relationship between the primary sequence of plant UGTs and acceptor substrate regioselectivity is not clear. For example, among the 11 Arabidopsis UGTs that catalyze glycosylation on the 3′-hydroxyl of quercetin, identity at the amino acid level is from 20 to 72%, and the betalain (betanidin) 6-O- and 5-O-glycosyltransferases from Dorotheanthus bellidiformis share only 19% amino acid sequence identity (25).
UGT71G1 is a relatively promiscuous UGT from the model legume *Medicago truncatula*. Its highest in vitro activity appears to be with quercetin, but it also efficiently catalyzes glycosylation of the isoflavone phytoestrogen genistein (5). Incubation of UGT71G1 with quercetin and UDP-glucose results in the formation of all five possible monoglucosides, with the 3′/H11032-O-glucoside predominating, whereas genistein is only glucosylated on the A-ring 7-hydroxyl (5, 30).

*EXPERIMENTAL PROCEDURES*

**Materials**—General chemicals were purchased from Sigma. UDP-[U-14C]glucose (300 mCi/mmol) was from American Radiolabeled Chemicals, Inc. (St. Louis, MO) and flavonoids and isoflavonoids were from Indofine Chemical Co., Inc. (Hillsborough, NJ).

**Site-directed Mutagenesis**—Site-directed mutagenesis was performed using the QuikChange II XL site-directed mutagenesis kit (Stratagene). The UGT71G1 cDNA (30) cloned into pET28a vector with a hexa-histidine tag (Novagen, Milwaukee, WI) was used as template. Synthetic oligonucleotides used for mutagenesis are listed in Table 1. Mutations were confirmed by sequencing.

**Expression of Recombinant Mutant Enzymes**—*Escherichia coli* BL21(DE3) cells harboring the wild-type and mutated constructs were cultured in LB medium at 37 °C until A600 reached 0.6–0.8. Isopropyl β-D-galactopyranoside was added to a final concentration of 0.5 mM and the cultures incubated overnight at 16 °C. UGT protein was purified using the MagneHis protein purification system according to the manufacturer’s instructions (Promega). pET28a-transformed *E. coli* BL21(DE3) cells were treated in parallel as a control. Protein concentration was determined with the Bio-Rad protein dye-binding assay (Bio-Rad) using bovine serum albumin as standard.

**Assay of Enzyme Activity**—Assays for determination of substrate specificity and regioselectivity were carried out in 50 mM Tris-HCl buffer (pH 7.0) containing 4 μg of purified recombinant protein, 500 μM UDPG, UDP-galactose, or UDP-glucuronic acid, 250 μM of either quercetin or genistein, and 14 mM mercaptoethanol in a final volume of 200 μl for 1 h at 30 °C. Reactions were stopped with 5 μl of trichloroacetic acid (240 mg/ml) and extracted with 250 μl of ethyl acetate. The extracted residues were resuspended in methanol and analyzed by reverse-phase HPLC (Hewlett Packard 1100 system) on a
5-μm C<sub>18</sub> column (Waters spherisorb 5-μm ODS2) with the following gradient: A = 1% phosphoric acid, B = acetonitrile, 0–5 min, 5% B; 5–10 min, 10% B; 10–25 min, 17% B; 25–30 min, 23% B; 30–65 min, 50%B; 65–69 min, 100% B; 69–79 min, 100% B; 79–80 min, 5% B. The eluate was monitored at 254 nm.

The glycosylated products were determined according to their relative retention times (27), UV spectra, and comparison to authentic standards. For kinetic studies, reaction mixtures contained 50 mM Tris-HCl (pH 7.0), 14 mM 2-mercaptoethanol, 10 μM UDP-[U-14C]glucose, 490 μM unlabeled UDPG, acceptor substrate (1.0–128 μM), and 2 μg of purified recombinant protein in a final volume of 50 μl. Reactions were incubated at 30 °C for 15 min, stopped by addition of 2 μl of trichloroacetic acid (240 mg/ml) and extracted with 200 μl of ethyl acetate. The organic phase was collected and radioactivity was determined by liquid scintillation counting (Beckman LS6500). Kinetic parameters were determined by hyperbolic regression analysis (Hyper32 program).

**RESULTS**

Effects of Mutation of Acceptor Binding Pocket Residues on Enzyme Activity and Regioselectivity for Quercetin—In *in vitro* assays with the flavonol quercetin as acceptor, UGT71G1 catalyzes formation of all five potential monoglucosides (30), with the 3-O-glucoside accounting for 70% of the total activity (Table 2, Fig. 2A).

To investigate the structural basis for the substrate specificity and product regioselectivity of UGT71G1, each of the amino acid side chains of the residues around the acceptor binding pocket were independently altered by site-directed mutagenesis.

**TABLE 1**

Primers used for introducing amino acid changes into UGT71G1

The modified bases are underlined.

| Mutation | Site<sup>a</sup> | Primer sequence |
|----------|-----------------|----------------|
| P18A     | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| I20V     | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| F49V     | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| G51R     | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| M52L     | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| P53A     | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| F54V     | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| L85R     | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| P88A     | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| E89K     | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| F148V    | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| Y202A    | A               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| S285A    | D               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| M286L    | D               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| W339G    | D               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| Q342A    | D               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| H357D    | D               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| N361A    | D               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| S362A    | D               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| E365P    | D               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |
| Y379A    | D               | CTCACTTCTTACCTCGCAACGAGAATGGCCACTTACGAGTAC |

<sup>a</sup> Mutation in sugar donor site (D) or phenolic acceptor site (A).

**TABLE 2**

Relative activities of wild-type and mutant UGT71G1 enzymes for glycosylation of the hydroxyl groups of quercetin with UDP-glucose as sugar donor

Results are expressed as HPLC peak area (average of three determinations) for each product as a % of total product peak area (%T), or as a % of the peak area of the same product from an identical incubation with an equal amount of wild-type enzyme (%W).

| Enzyme<sup>a</sup> | 5-OH | 7-OH | 3-OH | 4'-OH | 3'-OH | Total |
|---------------------|------|------|------|-------|-------|-------|
| 5-OH    | 7-OH | 3-OH | 4'-OH | 3'-OH |       |       |
| Wild-type | 8    | 100  | 2    | 100   | 7     | 100   | 8     | 100   |
| P18A     | 19   | 121  | 13   | 314   | 9     | 76    | 11    | 77    |
| I20V     | 8    | 55   | 2    | 48    | 7     | 60    | 10    | 74    |
| F49V     | 8    | 85   | 7    | 257   | 9     | 115   | 13    | 154   |
| G51R     | 10   | 85   | 2    | 63    | 8     | 84    | 9     | 88    |
| M52L     | 14   | 76   | 2    | 45    | 10    | 65    | 13    | 80    |
| P53A     | 10   | 92   | 4    | 124   | 8     | 90    | 12    | 115   |
| F54V     | 13   | 34   | 3    | 26    | 10    | 33    | 18    | 52    |
| L85E     | 10   | 50   | 1    | 23    | 7     | 40    | 8     | 43    |
| P88A     | 5    | 57   | 2    | 70    | 10    | 133   | 11    | 126   |
| E89K     | 0    | 0    | 0    | 0     | 0     | 0     | 0     | 0     |
| F148V    | 0    | 0    | 3    | 143   | 95    | 1374  | 1     | 13    |
| Y202A    | 0    | 0    | 3    | 143   | 95    | 1374  | 1     | 13    |
| M286L    | 4    | 47   | 1    | 48    | 3     | 46    | 1     | 15    |
| Y379A    | 12   | 27   | 26   | 229   | 4     | 12    | 5     | 14    |
| Vector control | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

<sup>a</sup> A, mutation in acceptor binding site; D, mutation in donor binding site.
TABLE 3
Kinetic parameters for wild-type and mutant (acceptor binding site) UGT71G1 enzymes with quercetin as acceptor

| Enzyme     | $V_{\text{max}}$ (μmol min$^{-1}$) | $K_m$ (μM) | $K_{\text{cat}}$ (s$^{-1}$) | $K_{\text{cat}}/K_m$ (s$^{-1}$ μM$^{-1}$) |
|------------|-----------------------------------|------------|-----------------------------|------------------------------------------|
| Wild type  | 0.81                              | 57.7       | 1.75 $\times$ 10$^{-2}$     | 303.3                                    |
| F148V      | 0.08                              | 19.5       | 1.66 $\times$ 10$^{-2}$     | 85.1                                     |
| Y202A      | 0.35                              | 27.1       | 7.5 $\times$ 10$^{-2}$      | 276.8                                    |

TABLE 4
Activity of wild-type and mutant (acceptor binding site) UGT71G1 enzymes with genistein and UDP-glucose

The activities shown as HPLC peak areas from identical incubations with equal amounts of protein, and the data are from 3 independent experiments.

| Enzyme     | Peak area | Activity relative to wild-type (%) |
|------------|-----------|-----------------------------------|
| Wild-type  | 8526      | 100                               |
| P18A       | 3825      | 44                                |
| I20V       | 5034      | 59                                |
| F49V       | 3453      | 41                                |
| G51R       | 5894      | 69                                |
| M52L       | 7499      | 87                                |
| P53A       | 8005      | 93                                |
| F54V       | 2666      | 31                                |
| L85R       | 8946      | 105                               |
| P88A       | 3202      | 37                                |
| E89K       | 0         | 0                                 |
| F148V      | 0         | 0                                 |
| Y202A      | 1453      | 17                                |

* Levels of putative genistein 5-O-glucoside.

sis (Fig. 1B, Table 1). Soluble enzyme expressed from wild-type and mutant enzyme constructs was obtained in yields from 4 to 24 μg/10 ml of induced culture, and SDS-PAGE analysis confirmed that a single band of the correct molecular weight was seen in the purified preparations of each mutant, suggesting that the single amino acid changes did not affect protein stability.

Reducing the size of large aromatic side chains at one end of the pocket dramatically altered regioselectivity with quercetin as acceptor, from predominantly B-ring 3'-O-glucosylation to predominantly C-ring 3-O-glucosylation (Table 2, Fig. 2A). Y202A retained 96% of wild-type activity with the 3-O-glucoside accounting for 95% of the conversion and production of the 3'-O-glucoside reduced to just 1% of the wild-type value (Table 2, Fig. 2A). Although the F148V mutant lost 80% of the wild-type activity, it produced a single product identified as quercetin 3-O-glucoside (Fig. 2A).

Kinetic analysis indicated that Y202A exhibited comparable catalytic efficiency ($K_{\text{cat}}/K_m$) to the wild-type enzyme, but that F148V exhibited 3.6-fold reduced catalytic efficiency (Table 3). The $K_m$ values for both mutant enzymes were 2–3-fold lower than for the wild-type enzyme, with the loss in efficiency of F148V resulting from a large decrease in $K_{\text{cat}}$.

The overall activities of the P18A, I20V, F49V, G51R, M52L, P53A, and P88A mutants were from 40 to 90% of wild-type activity. The activity of F54V was reduced by about 80% (Table 2). All these mutants preferentially glucosylated the 3'-hydroxyl of quercetin, with this reaction accounting for 48–76% of total products formed (Table 2). They also retained their ability to glucosylate the A-ring 5-hydroxyl of quercetin. Pro$^{18}$ and Phe$^{19}$ are located in the middle of the acceptor binding pocket, close to each other in the three-dimensional structure (Fig. 1). The P18A and F49V mutations caused a 2–3-fold increase in 7-O-glucosylation compared with the wild-type enzyme (Table 2). E89L had no detectable activity.

Effects of Mutation of Acceptor Binding Pocket Residues on Enzyme Activity and Regioselectivity for Genistein—Wild-type UGT71G1 only glucosylates the A-ring 7-hydroxyl of the isoflavone genistein (5, 30) (Fig. 2B), consistent with the orientation of genistein docked into the acceptor binding pocket (Fig. 1B). However, Y202A converted genistein to two products, the 7-O-glucoside and a compound tentatively identified as the 5-O-glucoside (comparison with an authentic standard ruled out production of the other possible product, the 4'-O-glucoside), with a total activity of 23% of that of the wild-type enzyme (Table 4, Fig. 2B). In contrast to its retention of activity with quercetin, F148V had no detectable activity with genistein (Table 4).

The P18A, I20V, F49V, G51R, M52L, P53A, F54V, and P88A mutants retained from 31 to 90% of the relative activities of the wild-type enzyme (Table 4), whereas L85R retained full activity. All these mutants produced only the 7-O-glucoside.

Effects of Mutation of Sugar Donor Binding Pocket Residues on Enzyme Activity and Sugar Donor Specificity—UGT71G1 displayed 1–4% activity with quercetin when UDPG was replaced by UDP-galactose or UDP-glucuronic acid. The 3'-O-glucoside was the predominant product (Fig. 2, C and D).

To investigate the roles of residues around the sugar binding pocket in sugar donor specificity and enzyme activity with quercetin as acceptor, site-directed mutagenesis was used to alter the side chains of those residues with close contacts to the sugar moiety (30) (Fig. 1D, Table 1). M286L retained full activity with UDP-glucose as sugar donor (Table 2). However, it exhibited 2–3-fold higher activity than the wild-type enzyme with UDP-galactose or UDP-glucuronic acid (Table 5). Y379A retained 65% of wild-type activity with UDP-galactose, but no activity was detected with UDP-glucuronic acid (Table 5). The S285A, W339G, Q342A, H357D, W360G, N361A, S362A, E365P, and Q382P mutants had no detectable activity with quercetin and UDP-glucose, UDP-galactose, or UDP-glucuronic acid (data not shown).

DISCUSSION

Based on the x-ray crystal structure of UGT71G1, a series of mutant enzymes with single alterations in amino acids surrounding the sugar donor and phenolic acceptor pockets were...
generated. Through enzyme activity assay with quercetin, residues Phe<sup>148</sup> and Tyr<sup>202</sup> were identified as crucial in determining the preferred 3′-O-regioslectivity of the wild-type enzyme. The Y202A mutant retained activity with the isoflavone genistein, but genistein 5-O-glucoside was formed in addition to the 7-O-glucoside, the only product formed by the wild-type enzyme. Structural analysis and molecular docking indicate that the 7-OH of genistein may be docked into the active site in a position to accept the glucose moiety from UDPG, but the proximity of Glu<sup>89</sup> and Ala<sup>380</sup> to the A-ring of genistein provide steric constraints for correct alignment of the 5-hydroxyl for glycosylation (Fig. 1B). Mutation of Tyr<sup>202</sup> to alanine creates more available space in the acceptor pocket and may allow the 5-OH to fit into the active site without these potential spatial conflicts.

With quercetin as acceptor, F148V only produced the 3-O-glucoside, and lost activity for other positions (including the 7-OH). The 7-OH of genistein can still be docked into the active site of the F148V mutant enzyme, and the reason for the loss of activity of this mutant is therefore not clear.

The E89K mutant lost activity with both genistein and quercetin. Glu<sup>89</sup> is located inside the acceptor binding pocket, close to the 5-OH or 7-OH groups of genistein docked into the active site. Although modeling shows that the E89K mutant might retain a similar overall acceptor binding pocket conformation, the pocket may be crowded due to the long side chain of lysine, and the change in electrostatic properties from negative to positive charge might negatively affect acceptor binding.

Quercetin is one of the major antioxidants in fruits and vegetables, and is consumed in forms in which one or more of its hydroxyl groups are glycosylated. The nature of the position of glycosylation significantly affects its absorption and utilization in humans (32–34). The chemical synthesis of specific quercetin glucosides is both complex and costly (35, 36). Recently, there has been increasing interest in biocatalytic synthesis of glucosides of quercetin (27, 37) and other phenolic compounds (38) due to their potential therapeutic value in cardiovascular disease and cancer, and their antimicrobial properties. Specific quercetin mono- or diglucosides have been successfully produced in an E. coli fermentation system employing regiospecific glycosyltransferases (27, 37).

The majority of the soluble glucoside products are secreted to the medium that facilitates their purification. We have now demonstrated that new glycosyltransferases with altered regioslectivity for quercetin can be obtained through targeted mutagenesis. Such mutant enzymes may prove useful in the biocatalytic synthesis of specific quercetin glucosides, because the position of conjugation affects both the bioavailability and health beneficial activity of quercetin for humans (20, 21). It remains to be seen whether the basic structure of UGT71G1 can be altered for preferential formation of the A-ring and B-ring (4′-O) glucosides.

Acknowledgments—We thank Drs. Luzia Modolo and Michael Udvardi for critical reading of the manuscript.

REFERENCES
1. Vogt, T., and Taylor, L. P. (1995) Plant Physiol. 108, 903–911
2. Miller, K. D., Guyon, V., Evans, J. N. S., Shuttleworth, W. A., and Taylor,
