Published in final edited form as:

Xenobiotica. 2012 August ; 42(8): 808–820. doi:10.3109/00498254.2012.663515.

Understanding Substrate Selectivity of Human UDP-glucuronosyltransferases through QSAR modeling and analysis of homologous enzymes

Dong Dong, Roland Ako, Ming Hu, and Baojian Wu

Department of Pharmacological and Pharmaceutical Sciences, College of Pharmacy, University of Houston, Houston, TX 77030, USA

Abstract

1. The UDP-glucuronosyltransferase (UGT) enzyme catalyzes the glucuronidation reaction which is a major metabolic and detoxification pathway in humans. Understanding the mechanisms for substrate recognition by UGT assumes great importance in an attempt to predict its contribution to xenobiotic/drug disposition *in vivo*.
2. Spurred on by this interest, 2D/3D-quantitative structure activity relationships (QSAR) and pharmacophore models have been established in the absence of a complete mammalian UGT crystal structure.
3. This review discusses the recent progress in modeling human UGT substrates including those with multiple sites of glucuronidation. A better understanding of UGT active site contributing to substrate selectivity (and regioselectivity) from the homologous enzymes (i.e., plant and bacterial UGTs, all belong to family 1 of glycosyltransferase (GT1)) is also highlighted, as these enzymes share a common catalytic mechanism and/or overlapping substrate selectivity.

Keywords

Glucuronidation; UGTs; Crystal structure; *In Silico* modeling; QSAR; Pharmacophore; CoMFA

Introduction

Glucuronidation catalyzed by UDP-glucuronosyltransferases (UGTs) is the principal phase II metabolism, whereby the glucuronic acid (from the cofactor UDP-glucuronic acid) is conjugated to the substrate containing a nucleophilic group (usually the hydroxyl, carboxylate, or amines). It has been recognized that serum bilirubin (the neurotoxic product of hemoglobin) is mainly cleared via glucuronidation; whose deficiency could result in jaundice and a disease state known as hyperbilirubinemia (Bosma, 2003). Human UGTs are classified into four families: UGT1, UGT2, UGT3, and UGT8, on the basis of amino acid

sequence identity (Mackenzie et al., 2005). The most important drug-conjugating UGTs belong to UGT1 and UGT2 subfamilies or alternatively UGT1A, UGT2A, and UGT2B subfamilies. The UGT1A and 2B isoforms are abundantly expressed in the intestine and liver, two major metabolic organs. These isoforms mediate intestinal and hepatic (first-pass) glucuronidation, limiting the oral bioavailability of many phenolics including drugs and natural phenols (Wu et al., 2011a).

Due to the reaction nature (a nucleophilic substitution reaction), UGT isoforms generally accept a wide variety of different types (structurally unrelated) of substrates, including drugs (e.g., raloxifene and mycophenolic acid), dietary chemicals (e.g., hydroxycinnamic acids and flavonoids), environmental toxins (e.g., benzo(a)pyrene and nitrosamines), endogenous compounds (e.g., bile acids, and steroid hormones). In addition, one substrate is usually glucuronidated by several isoforms (i.e., overlapping substrate specificity). This promiscuity of broad and overlapping substrate selectivity is advantageous for enzymes involved in detoxification, but it poses significant challenges in identifying a selective UGT probe and in understanding of the mechanisms determining the substrate selectivity (Court, 2005).

A complete three dimensional structure of human UGTs is not yet available. Only a partial crystal structure of UGT2B7 (C-terminal domain) was resolved (Miley et al., 2007). This structure, combined with molecular modeling studies, provides substantial insights into the UDPGA binding and catalytic mechanism (Miley et al., 2007; Radomska-Pandya et al., 2010). Due to the lack of structural information of the N-terminus (for substrate binding), understanding the substrate selectivity of human UGTs is still very difficult. Investigators have employed various 2D/3D-QSAR techniques and pharmacophores to characterize the chemical features for UGT substrates, and to generate robust models for high throughput prediction of metabolism (Smith et al., 2004; Sorich et al., 2008). The contribution of these *in silico* models to understand the UGT active site is discussed in this review. Careful analyses of the crystal structures and substrate selectivity (including regioselectivity) for the homologous enzymes (i.e., plant and bacterial UGTs) are also provided. Please note that substrate selectivity here refers to the recognition of a sugar acceptor not a UDP-sugar donor. High similarities in substrate selectivity are found between human UGTs and plant UGTs. This leads to an improve understanding of human UGTs using structure-function correlations derived for plant UGTs. Please note that the term “UGT” if used alone stands for human UGTs throughout this article.

Glycosyltransferase (GT) 1 enzymes

Glycosyltransferases (GTs) catalyze the glycosylation reaction by transferring a sugar residue derived from the cofactor (nucleotide sugar or sugar donor) to the acceptor substrates. The GTs constitute a large number of enzymes, which have been classified into 94 families (<http://www.cazy.org>) on the basis of primary sequence identity. Despite low sequence identity, GTs demonstrate a high conservation in their 3D structures; they adopt either GT-A fold or GT-B fold (Hu and Walker, 2002; Zhang et al., 2003; Breton et al., 2006; Lairson et al., 2007). The GT-A fold consists of two closely abutting $\beta/\alpha/\beta$ Rossmann domains, whereas the GT-B fold consists of two $\beta/\alpha/\beta$ Rossmann domains that face each other and are linked flexibly with the active site located at domain interface (Figure 1A).

Most GT-A enzymes possess a DxD motif, which is involved in coordinating a divalent cation (typically Mn^{2+}) in the catalytic center. The metal ion is required for the binding of the nucleotide sugar. By contrast, although divalent cations (e.g., Mg^{2+}) may be required for full activity of GT-B enzymes, there is no evidence of a metal ion bound in the GT-B structures (Breton et al., 2006). The mechanism of GT-mediated glycosylation is defined as either inverting or retaining (Figure 1B) (Coutinho et al., 2003). Human UGTs belong to the GT1 family. Enzymes of this family adopt the GT-B fold and follow an inverting mechanism (Osmani, et al., 2009; Radomska-Pandya et al., 2010). By far, the GT1 family contains five full 3D structures from plants and ten from bacteria (Table 1). Four of the plant GT1 structures, UGT71G1, UGT78G1, UGT85H2, and V_{ν} GT1 are flavonoid glucosyltransferases (Shao, et al., 2005; Offen, et al., 2006; Li, et al., 2007a; Modolo et al., 2009), and UGT72B1 is a bifunctional N- and O-glucosyltransferase (Brazier-Hicks, et al., 2007a).

Among the ten bacterial structures, TDP-*epi*-vancosaminyltransferase GtfA, UDP-glucosyltransferase GtfB and TDP-vancosaminyltransferase GtfD are involved in vancomycin synthesis (Mulichak, et al., 2001, 2003, 2004); four calicheamicin glycosyltransferases (CalG1, CalG2, CalG3, and CalG4) are involved in calicheamicin synthesis (Zhang, et al., 2008; Chang et al., 2011); the oleandomycin glycosyltransferases OleD and OleI inactivate oleandomycin and diverse macrolide antibiotics (Bolam, et al., 2007); and glycosyltransferase UrdGT2 catalyses the formation of a C-C bond between a polyketide aglycone and D-olivose (Mittler, et al., 2007). Although the sequence identity between human UGTs and those GT1 enzymes whose structures have been solved is typically below 20%, the structural fold is predicted to be highly conserved (Coutinho et al., 2003; Breton et al., 2006).

Catalytic mechanisms for human UGTs

It is becoming evident that the catalytic mechanisms for human UGTs are highly complex. Human UGTs use distinct catalytic mechanisms for O-glucuronidation vs. N-glucuronidation. The catalytic mechanism for O-glucuronidation is described as the “serine hydrolase like mechanism”, which involves two key amino acids histidine and aspartic acid (so-called “catalytic dyad” or “acid base pair”) (Battaglia et al., 1994; Miley et al., 2007; Li et al., 2007b; Patana et al., 2008). In this mechanism, the catalytic histidine (base) abstracts a proton from the aglycone hydroxyl for nucleophilic attack at the C1 carbon of glucuronic acid in UDPGA. The proton abstraction to the histidine is stabilized by the neighboring aspartate (acid) (Figure 2A). Chemically, this deprotonation process is necessary to avoid the enormous barrier to the formation of the positively charged oxonium-ion [$R_1-OH^+-R_2$, $pK_a \approx -4$] (Brazier-Hicks, et al., 2007a). Consistent with their catalytic role, mutations of the His-Aps dyad in UGT1A1, 1A9, and 2B7 abolish or markedly reduce the enzyme activity (Li et al., 2007b; Miley et al., 2007; Patana et al., 2008). Interestingly, many other GT1 enzymes including plant UGTs (UGT71G1, UGT72B1, UGT82H5, UGT78G1, and V_{ν} GT1) and bacterial UGTs (CalG1, CalG3, CalG4, OleD, and OleI) also use this mechanism for O-catalysis (Figure 3). However, it is also noted that the GT1 enzyme isoflavone 7-O-glucosyltransferase may utilize a strategy that is different from the serine

hydrolase like mechanism for O-catalysis, because the corresponding “catalytic dyad” His-15 and Asp-125 are not important for enzyme activity (Noguchi et al., 2007).

It has been recognized that the catalytic dyad for GT1 enzymes is not limited to the His-Asp. The catalytic base may also be an aspartate (e.g., GtfA, GtfA, GtfB) or a threonine (e.g., UrdGT2), though a corresponding amino acid for stabilizing effect is unknown (Mulichak et al., 2001, 2003, 2004). The catalytic acid may be the glutamate (e.g., CalG1 and CalG3), which hydrogen bonds to the ND1 atom of histidine. However, in the GtfB, the putative base D13A mutant retained ~10% activity, leading the authors to suggest that an alternative catalytic base (Asp332) is located in the C-, rather than the N-terminal domain. Also, the ability of aspartate in UGT1A9 to be a catalytic base is questioned, because the H37D mutation markedly decreased the enzyme activity (Patana et al., 2008). Moreover, the magnitude of reduction in enzyme activity caused by the H37D mutation is more significant than that by the H37A mutation (Patana et al., 2008).

There is compelling evidence that N-glucuronidation uses a different mechanism that does not require proton abstraction of the substrate, compared to O-glucuronidation (Figure 2B). In the absence of deprotonation, N-nucleophiles can still develop a charged secondary amine [$R_1-NH_2^+-R_2$, $pK_a \approx 5$] as the transition state, which is a stable and readily attained species in contrast to an oxonium-ion (Brazier-Hicks et al., 2007a). Therefore, it is not surprising that UGT1A4 and UGT2B10, the only two human UGTs that lack the “catalytic” histidine, completely lose the O-glucuronidation activity and are more specialized for N-glucuronidation (Kerdpin et al., 2009; Kaivosaaari et al., 2011). Patana et al (2008) propose that a negatively charged residue (most likely an aspartic acid) in the UGT1A9 protein is required to stabilize the positively charged amine in the transition state, based on the observation that mutations of two aspartic acids (Asp143 and Asp148) leads to reduced N-glucuronidation activity. However, at present, whether N-glucuronidation requires the participation of a negatively charged residue remains uncertain.

Several human UGT isoforms such as UGT1A9 are capable of catalyzing both O-glucuronidation and N-glucuronidation (Patana et al., 2008). This raises the questions whether and how the “catalytic” histidine (His39) contributes to N-glucuronidation. The influence of the catalytic histidine on N-catalysis may be significant because the UGT1A9 mutant (H39A)-mediated glucuronidation shows marked decreases in the V_{max} value (Patana et al., 2008). Brazier-Hicks et al. (2007a) propose the possible role of the “catalytic” histidine in N-glucuronidation of aniline-type substrates. In their proposal, the histidine functions to direct and orientate nucleophilic attack. Specifically, the histidine hydrogen-bonds to a long pair on the substrate to prevent conjugation of the amine long pair with the aromatic system. This would increase the electron density of the amine group and the nucleophilicity of the substrate (Brazier-Hicks et al., 2007a). Interestingly, in a crystal structure for plant UGT72B1 (a bifunctional O- and N-glucosyltransferase), the enzyme displays a non-canonical GT1 geometry and uncoupling of the catalytic dyad (His-Asp) (Brazier-Hicks et al., 2007a). The authors believe that this conformation observed is used for N-catalysis (Brazier-Hicks et al., 2007a). Therefore, it may reveal the GT1 enzyme’s adaptability to catalyze substrates with different types of nucleophilic group using distinct mechanisms (so-called “substrate-defined catalysis”) (Patana et al., 2008).

In silico models for human UGTs

In silico methods are widely used to elucidate the molecular basis for substrate recognition by drug-metabolizing enzymes (Ekins et al., 2001). In the absence of a human UGT crystal structure, the 2D/3D-QSAR and pharmacophore techniques have been applied to establish predictive models for UGT isoforms, and to derive the chemical features for their substrates since early 2000.

UGT1A1

UGT1A1 perhaps is one of the most important enzymes in the literature due to its major role in clearance of bilirubin and drugs (e.g., SN-38). The UGT1A1 polymorphism UGT1A1*28 has been associated with the hyperbilirubinemia disorder and the toxicity of SN-38 (O'Dwyer and Catalano, 2006). The only UGT1A1 QSAR study was performed with 23 structurally diverse substrates, using the apparent inhibition constant ($K_{i,app}$) as the specific activity (Sorich et al., 2002). $K_{i,app}$ was derived using 4-methylumbelliferone as the 'probe' substrate, assuming by the authors that the inhibition data follow the competitive inhibition model. The whole dataset was divided into the training set ($n = 18$) and test set ($n = 5$); the former was used for model construction and the latter for model validation (Sorich et al., 2002). This study produces three different models with good predictive power, namely, a 2D-QSAR model ($r^2 = 0.92$, $K_{i,app}$ for all five test substrates predicted within 1 log unit), a 3D-QSAR (SOMFA) model ($r^2 = 0.71$, $K_{i,app}$ for four out of the five test substrates predicted within 1 log unit), and a (discriminative features) pharmacophore model ($r^2 = 0.87$, $K_{i,app}$ for all five test substrates predicted within 1 log unit).

The 2D-QSAR model includes a descriptor logP, suggesting that substrate hydrophobicity is a determinant of substrate binding to the active site of UGT1A1 (Sorich et al., 2002). This is consistent with the fact that UGT active site is located on the luminal site of microsomal membrane. However, interpretation of other descriptors from the model is very difficult (Sorich et al., 2002). The discriminative features pharmacophore indicates that an aromatic ring, a hydrogen bond donor, and a hydrophobic region are important for substrate binding. By contrast, the common features pharmacophore indicates that two hydrophobic domains on the substrate are commonly found close to the glucuronidation site of the chemical (Figure 4).

UGT1A3

There is no QSAR study about UGT1A3 in the literature. Little is known about the structure features for UGT1A3 substrates. Qualitative analysis of the UGT1A3 activity against a series of flavonoids revealed that the preferred substrates contain the hydroxyl group at the C7-position, mostly likely is the site of glucuronidation (Wu et al., 2011b; Xie et al., 2011).

UGT1A4

UGT1A4 catalyzes glucuronidation of amine groups (N-glucuronidation). The first UGT1A4 QSAR was performed using eight 1-substituted imidazoles (Vashishtha et al., 2001). Linear correlations are found between V_{max} and both the partition coefficient (logP) and pKa, and between K_m and pKa, indicating that the lipophilicity and the availability of

long pair of electrons on the substrate are important for catalysis. However, the authors acknowledge that these correlations should be treated with caution, due to the low number of substrates examined and limited range of values for each glucuronidation parameter. Smith et al. (2003a) published one 2D-QSAR and two 3D-QSAR models (i.e., a pharmacophore based and a molecular field based model) for UGT1A4 using 24 structurally diverse substrates. The 24 molecules were divided into training ($n=18$) and test set ($n=6$); the former was used for model construction and the latter for model validation (Smith et al., 2003a). All six molecules in the test set (not included in model derivation) were predicted within 1 log unit using either derived models, suggesting the models can be of great value in prediction of UGT1A4 metabolism. The 2D-QSAR exhibits the best overall predictability; the inclusion of logP in the model indicates the substrate hydrophobicity is important in substrate-enzyme interaction. However, qualitative interpretation of all other descriptors is difficult. Interestingly, both 3D-QSAR and common features pharmacophore models include two hydrophobic regions in addition to the glucuronidation site (Figure 4).

UGT1A6

Ethell et al. (2002) reported the first UGT1A6 QSAR ($r^2 = 0.996$; $q^2 = 0.98$) using 10 simple 4-substituted phenols. This model successfully relates the log K_m values to the molecular surface (MS-WHIM) and atomic descriptors (AT-WHIM), demonstrating that a QSAR method can be applied to analyze glucuronidation data. However, due to the narrow range of K_m values (a 7-fold variation), correlation of the model features to the properties of the active site is limited. Interestingly, UGT1A6 V_{max} values decrease as the bulk of the substituent increases. This leads to the hypothesis that bulkier substrates may be accepted into the active site but may not be glucuronidated (Ethell et al., 2002). This is evidenced by the fact that the phenols with large 4-substituents, which are not UGT1A6 substrates, could inhibit 4-ethylphenol glucuronidation.

Sorich et al. (2004) developed a binary classification model which is a linear combination of three pharmacophores. This model was capable of predicting 78% of the test set of 58 chemicals correctly. Non-substrates in the test set were predicted slightly better than the substrates (80% vs. 72%). The first pharmacophore (P1), which was present in more substrates than non-substrates, described a nucleophile approximately 2 Å from an aromatic ring. This simple pharmacophore generally mapped to chemicals where the nucleophiles was directly attached to an aromatic ring, such as phenolic or aniline like functional groups. The second pharmacophore (P2) described two hydrophobic regions, one close to the nucleophile feature (2 Å) and the other distant to the nucleophile feature (6 Å). The presence of this pharmacophore in a chemical lowered the probability that it would be glucuronidated by UGT1A6. The final pharmacophore (P3) was also associated with non-substrates and contained hydrophilic and hydrophobic regions 6 and 8 Å from a nucleophile, respectively.

UGT1A7

The binary classification model published by Sorich et al (2004) is the only QSAR for UGT1A7 in the literature. This model consists of three pharmacophores. The predictability of the model was validated against the test set of 20 chemicals; 85% were predicted correctly. Non-substrates in the test set were predicted better than the substrates (92% vs.

72%). The first pharmacophore (P1) indicated that substrates tended to have a nucleophile attached to a hydrophobic aromatic ring. The second pharmacophore (P2) represented a nucleophile approximately 2 Å from a hydrogen bond acceptor. This pharmacophore recognized carboxylic acid and imidazole groups, both of which were more common in non-substrates. The third pharmacophore (P3) comprised a hydrogen bond donor approximately 8 Å from a nucleophile. This pharmacophore was also associated with non-substrates and was often detected in hydroxysteroids.

UGT1A9

The first QSAR study for UGT1A9 was performed using the K_m values for glucuronidation of 24 simple 4-substituted phenols (Ethell et al., 2002). The resulting model shows good internal and external consistency ($r^2 = 0.83$ and $q^2 = 0.73$). However, the authors acknowledge the limitation of the dataset (i.e., a narrow range of K_m values), and believe that their model is a good starting point for building a more complex model. In a later study, a more structurally diverse dataset (K_i values) of UGT1A9 was assembled for QSAR analyses (Smith et al., 2004). Unfortunately, no satisfactorily predictive model could be developed. Nonetheless, the identified common-features pharmacophore for UGT1A9 shares the core features (i.e., a glucuronidation site with two adjacent hydrophobic regions) with those for UGT1A1 and UGT1A4, which may help explain the overlapping substrate specificity between UGT1A family isoforms (Figure 4) (Sorich et al., 2002; Smith et al., 2003a,b; 2004). It is noted that a hydrogen-bond acceptor (HBA) (near the outmost hydrophobic region) is also a common feature for UGT1A9 substrates but not for UGT1A1 or UGT1A4 substrates (Sorich et al., 2002; Smith et al., 2003a,b; 2004).

In the study of Sorich et al. (2004), a binary classification model, namely, a linear combination of four pharmacophores, was successfully established to predict UGT1A9 substrates and non-substrates. The test set of 64 chemicals was predicted with 80% accuracy by the model. The first pharmacophore (P1) comprised a nucleophile attached to an aromatic ring. P1 was strongly associated with known substrates and represented functional groups such as phenol and aromatic amine. Additionally, an aromatic ring approximately 8 Å from a nucleophile (P2) was also associated with UGT1A9 substrates. Substrates commonly mapped to both P1 and P2. There were also two pharmacophores associated with non-substrates. Five times more non-substrates than substrates contained the configuration of two hydrophobes and a hydrogen bond acceptor all approximately 3 Å from the nucleophile (P3). This pharmacophore was typically associated with fused aromatic rings with the hydrogen bond acceptor situated adjacent to the nucleophile. The fourth pharmacophore (P4) was a hydrogen bond donor approximately 6 Å from a nucleophile. This pharmacophore was not obviously associated with any specific chemical substructure.

Wu et al. (2011c) reported two pharmacophore-based comparative molecular field analysis (CoMFA) models for UGT1A9-mediated glucuronidation of 30 flavonols using V_{max} and CL_{int} values, respectively. The models are position specific as they only account for glucuronidation at 3-OH group. The authors hypothesized that a predictive algorithm should consider more than one active binding mode because many UGT substrates possess multi-sites of glucuronidation for which more than one active binding mode are possible. This

hypothesis was validated in a following study (Wu et al., 2012). The authors constructed two generalized models for UGT1A9 using CoMFA and CoMSIA (comparative molecular similarity indices analysis) techniques based on a large number of phenolic substrates ($n = 145$). Multiple binding modes (or conformations) for each substrate with multi-sites of glucuronidation are included in the model construction and prediction. The resulting models are statistically significant with good predictive power (CoMFA: $q^2 = 0.548$, $r^2 = 0.949$, $r^2_{\text{pred}} = 0.775$; CoMSIA: $q^2 = 0.579$, $r^2 = 0.876$, $r^2_{\text{pred}} = 0.700$). This in turn strongly supports the “multiple binding modes” theory (Wu et al., 2012).

Moreover, the authors matched the CoMFA/CoMSIA maps with the active site from a homology model of UGT1A9 (Wu et al., 2012) (Figure 5A). The steric maps show a good compatibility with the active site of the homolog model (Figure 5B). The green areas (bulky group favorable) appear in the active site cavities where no amino acids are found. The yellow areas (bulky group unfavorable) take place in the regions occupied by the amino acids. Therefore, the steric interactions are highlighted between UGT1A9 substrates and the active site. A good UGT1A9 substrate should closely fit into the enzyme's active site for maximal steric interactions.

UGT1A10

To date, there has been no published UGT1A10 QSAR. Based on direct comparisons of the determined activities for bioflavonoids studied, Lewinsky et al (2005) proposed the features of the bioflavonoid structure necessary to confer it as a substrate of UGT1A10. The preferred substrates of UGT1A10 contain the hydroxyl group to be glucuronidated at C6 or C7, but not C5 of the A-ring or on C4' of the B-ring. Up to two additional hydroxyl groups on the A-ring enhance activity, whereas the presence of other groups, notably sugar groups, decreases activity.

UGT2B4

A binary model predicting the UGT2B4 substrate and non-substrate is available (Sorich et al., 2004). This model consisting of two pharmacophores predicted 86% of the 42 chemical test-set correctly, with substrates predicted more accurately than non-substrates (92 vs. 83%). The first pharmacophore (P1) was a nucleophile near (2–4 Å) three hydrophobic regions. This indicated that the site of glucuronidation was commonly situated near hydrophobic regions of the molecule. The second pharmacophore (P2) described a symmetrical arrangement between a hydrophobic region, a nucleophile, and a hydrogen bond acceptor. The nucleophile and hydrogen bond acceptor were separated by approximately 10 Å and each had a hydrophobic region nearby (2 or 3 Å). The symmetry of the pharmacophore often allowed individual chemicals map to the pharmacophore in two distinct ways.

Correlations of the substrate selectivity with the active site for plant UGT and bacterial UGT (the homologs of human UGT)

The substrate profiles for the five plant UGTs demonstrate that UGT72B1 accepts smaller molecule substrates compared to UGT71G1, UGT78G1, UGT85H2, or VvGT1 (Table 2).

The typical substrates of UGT72B1 are 2,4,5-trichlorophenol and 3,4-dichloroaniline. This substrate selectivity is consistent with the closed, small, and compact activity site of this enzyme (Brazier-Hicks et al., 2007a). The volume of UGT72B1 active site is estimated at 41 Å³, much smaller than those of UGT71G1, UGT78G1, UGT85H2, or VvGT1 (Figure 6). Please note that estimates of the active site volumes for the latter are inaccurate (greatly less than the actual values) as the active sites are in an opened form (open to the solvent). UGT71G1, UGT78G1, UGT85H2, and VvGT1 accept broad-ranging substrates such as plant hormones, flavonoids, coumarins, phenylpropanoids and benzoic acids. The bacterial UGTs (GtfA, GtfB, GtfD, CalG1, CalG2, CalG3, CalG4, OleD, OleI, and UrdGT2) involve in biosynthesis of antibiotics such as vancomycin and calicheamicin, and generally display exclusivity in the substrate selection.

Comparisons of the active site vs. the typical substrate between VvGT1 and UGT72B1 indicate that their substrate selectivity differences are highly determined by the active site properties (e.g., the size and shape) (Figure 7A). In order to accommodate a larger molecule (e.g., kaempferol vs. 2,4,5-trichlorophenol), the VvGT1 active site is significantly enlarged by the shifting of helix Na3 and Na5a (the naming is consistent with Osmani et al. (2009)) away from the active site core. Major structural differences in the active site (particularly in the region of Na5-Na5b) are obvious between plant VvGT1 and bacterial OleI. Accommodation of the large molecular oleandomycin requires not only the shift of secondary structural element (e.g., Na3) but also the disruption of secondary structures (e.g., Na5, Na5a, and Na5b) in enzyme OleI (Figure 7B).

It is generally favorable that the human UGT structure may share more similarities with the plant UGTs than bacterial UGTs because (1) the C-terminal structure of human UGT2B7 agrees well with those of plant UGTs (Miley et al., 2007); (2) human UGTs share overlapping substrate selectivity (e.g., coumarins and flavonoids) with plant UGTs; and (3) both human and plant UGTs demonstrate flexibility in substrate selection, whereas bacterial UGTs only acts on limited type of antibiotics. The wide substrate selectivity for plant UGTs is supported by the fact that the side chains of the residues forming the active site undergo conformation changes, even though little or none changes occur to the backbone structure (Figure 8). This flexibility in enzyme conformation for fitting distinct substrates has been frequently seen in drug-metabolizing enzymes such as cytochrome p450 and sulfotransferase (Berger et al., 2011; Dong and Wu, 2012).

On the other hand, it is noteworthy that fundamental differences between human and plant UGT exist. First, human UGT enzyme is membrane bound protein whereas plant UGT is soluble. Second, human UGT mainly uses UDP-glucuronic acid as the cofactor whereas plant UGT most commonly uses UDP-glucose, though these two sugar donors are highly structurally similar with only a carbonyl group difference. Interestingly, in some cases, plant UGT can also use UDP-glucuronic acid as the sugar donor (Osmani et al., 2009).

Broad vs. limited vs. overlapping substrate selectivities among human UGTs

Human UGT1A1 is known to accept a wide range of substrates, including bilirubin, simple phenols, coumarins, fatty acid, flavonoids, and steroid hormone. The common features pharmacophore model reveals that a UGT1A1 substrate should bear two hydrophobic regions in addition to one glucuronidation site (Sorich et al., 2002). These features are readily found in many classes of chemicals and drugs, thus it helps explain why a broad substrate selectivity is usually displayed by UGT1A1. By analogy to the plant UGTs, this broad substrate selectivity may be attributable to an open active site and the conformation flexibility in the side chains of the residues forming the active site, though the experimental evidence is still lacking.

In contrast, human UGT1A6 shows relatively strict substrate selectivity and has a preference for metabolizing small molecules such as simple phenols (Ethell et al. 2002). This is supported by the UGT1A6 substrate model, a nucleophile approximately 2 Å from an aromatic ring, which is perfectly mapped to phenols (Sorich et al., 2004). A possible mechanism proposed here is that UGT1A6 has a relatively small and closed active site, somewhat similar to plant UGT72B1 (Brazier-Hicks et al., 2007a). It is acknowledged that the validity of this proposal awaits a crystal structure for this enzyme.

Another important feature for human UGTs is that the enzymes display vastly overlapping substrate selectivity. It is interesting to note that the pharmacophore models for UGT1A1, 1A4, and 1A9 reveal a high similarity, demonstrating that these enzymes can act on a same substrate (Figure 4). Again, by utilizing the structure-function relationships for plant UGTs, the overlapping substrate selectivity among these human UGT enzymes may be resulted from a highly related active site, as observed for plant UGT78G1 and VvGT1 (Figure 6).

A hypothetical binding model to explain regioselectivity of human UGT1A1

Extensive kinetic characterization has generalized the regioselectivity for human UGT1A1 towards flavonoids (e.g., kaempferol and quercetin), 3'-OH > 7-OH > 3-OH (Wu et al., 2011b,d). However, the molecular basis for this strict regioselectivity remains unknown due to the lack of a crystal structure for this enzyme. By contrast, the substrate regioselectivity of plant UGTs are relatively well clarified owing to the solved crystal structures (Osmani et al., 2009). Formation of multiple glucuronides from a flavonol is because the flavonol molecule can fit in the active site with different orientations enabling the glycosylation of either hydroxyl group (Osmani et al., 2009). Plant UGT71G1 is the enzyme showing high selectivity of quercetin 3'-OH, which is favored by hydrogen-bond formation to S285 as well as to the sugar donor UDP-glucose (Shao et al., 2005). Positioning of 3-hydroxy group is not easy due to the steric effects of large aromatic residue F148 from the helix Nα5 and possible F202 from Nα5b (He et al., 2006). Due to lack of a bulky residue in the corresponding position, plant UGT78G1, UGT85H2 and VvGT1 show a high preference on the 3-OH position other than 3'-OH as observed for UGT71G1. The structural basis underlying the regioselective glycosylation for plant UGTs is also illustrated in Figures 6 & 9A.

With molecular evidence from plant UGTs, a class of homologous enzymes for human UGTs, a hypothetical binding model of flavonoids in the UGT1A1 active site could be deduced and is shown in Figure 9B using quercetin as an example. The inner part (or the end) of the substrate-binding pocket provides steric bulks (possible aromatic rings), which disfavor the positioning of quercetin for 3-OH catalysis (Figure 9B). Compared to 7-OH, 3'-OH is more preferred by the potential hydrogen bonds formation between 4'-OH and the binding pocket (Figure 9B). 4'-OH-mediated hydrogen bond is partially supported by the fact that glucuronidation of 3'-hydroxyflavone is markedly reduced by 5-folds in the absence of 4'-OH, compared to quercetin (unpublished data).

Conclusion

Glucuronidation has been increasingly recognized as an important reaction for metabolism and detoxification of numerous endogenous and exogenous compounds (including drugs). Despite its significance, relatively little is known about the UGT active site and how it interacts with the substrates. In last decade, various modeling techniques such as 2D/3D-QSAR (including molecular field analysis) and pharmacophores were employed to elucidate substrate selectivity of UGTs. A range of models were produced covering the naïve regression models independent of molecular conformation (interpretation of which usually is difficult), reaction site fortified pharmacophore models, and the sophisticated models accounting for multiple binding modes. It is encouraging that the models possess a high predictability, some of which were based a large number of structurally diverse substrates.

We believe that a more useful and reasonable UGT model should address the issues raised here. First, majority of UGT models in the literature seek for correlations between the binding affinity (K_m or K_i values) and the substrate structures, assuming binding affinity is a measure of the glucuronidation activity. However, this might not be always true. Poor or no correlation of K_m values with V_{max} or CL_{int} values has been observed for UGT1A9-mediated glucuronidation (Wu et al., 2011c). In other words, a poorly bound substrate is not necessarily associated with a low metabolic turnover, nor the vice versa (Ethell et al., 2002). Second, kinetic characterization on UGTs with the substrate with multi-sites of glucuronidation reveals the complexity of UGT-substrate interaction, indicating that UGT active site may permit multiple (active) binding modes (corresponding to glucuronidation at different sites) for a substrate (Wu et al., 2011b,d). Thus a predictive algorithm that assumes only one (active) binding mode for one substrate would be limited of use in ultimate elucidating substrate selectivity (and regioselectivity) of UGTs. A novel algorithm incorporating this "multiple binding modes" knowledge appears to be essential for a quality model.

Interestingly, significant commonalities such as the catalytic mechanism and substrate selectivity (including regioselectivity) are found between human UGTs and plant UGTs. Analogical analyses in this article demonstrate the structure-function correlations from plant UGTs may help understand the complex behaviors of human UGTs in substrate selection. Therefore, in the absence of a human UGT crystal structure, we anticipate that a close integration of plant UGTs with solved structures and molecular modeling techniques would provide a more robust model that is also intuitive and interpretable.

Acknowledgement

The authors are grateful to Dr. Diana Chow for financial support during this work. We also thank the reviewers' valuable suggestion and comments.

Abbreviations used

GTs	Glycosyltransferases
PDB	Protein Data Bank
QSAR	Quantitative Structure-Activity Relationship
UGTs	UDP-glucuronosyltransferases

References

- Battaglia E, Senay C, Fournel-Gigleux S, Herber R, Siest G, Magdalou J. The chemical modification of human liver UDP-glucuronosyltransferase UGT1*6 reveals the involvement of a carboxyl group in catalysis. *FEBS Lett.* 1994; 346(2–3):146–150. [PubMed: 8013623]
- Berger I, Guttman C, Amar D, Zarivach R, Aharoni A. The molecular basis for the broad substrate specificity of human sulfotransferase 1A1. *PLoS One.* 2011; 6(11):e26794. [PubMed: 22069470]
- Bolam DN, Roberts S, Proctor MR, Turkenburg JP, Dodson EJ, Martinez-Fleites C, Yang M, Davis BG, Davies GJ, Gilbert HJ. The crystal structure of two macrolide glycosyltransferases provides a blueprint for host cell antibiotic immunity. *Proc Natl Acad Sci U S A.* 2007; 104(13):5336–5341. [PubMed: 17376874]
- Bosma PJ. Inherited disorders of bilirubin metabolism. *J Hepatol.* 2003; 38:107–117. [PubMed: 12480568]
- Strassburg CP. Pharmacogenetics of Gilbert's syndrome. *Pharmacogenomics.* 2008; 9:703–715. [PubMed: 18518849]
- Brazier-Hicks M, Edwards LA, Edwards R. Selection of plants for roles in phytoremediation: the importance of glucosylation. *Plant Biotechnol J.* 2007b; 5(5):627–635. [PubMed: 17573805]
- Brazier-Hicks M, Edwards R. Functional importance of the family 1 glucosyltransferase UGT72B1 in the metabolism of xenobiotics in *Arabidopsis thaliana*. *Plant J.* 2005; 42(4):556–566. [PubMed: 15860014]
- Brazier-Hicks M, Offen WA, Gershater MC, Revett TJ, Lim EK, Bowles DJ, Davies GJ, Edwards R. Characterization and engineering of the bifunctional N- and O-glucosyltransferase involved in xenobiotic metabolism in plants. *Proc Natl Acad Sci U S A.* 2007a; 104(51):20238–20243. [PubMed: 18077347]
- Breton C, Snajdrová L, Jeanneau C, Koca J, Imberty A. Structures and mechanisms of glycosyltransferases. *Glycobiology.* 2006; 16(2):29R–37R. [PubMed: 16049187]
- Chang A, Singh S, Helmich KE, Goff RD, Bingman CA, Thorson JS, Phillips GN Jr. Complete set of glycosyltransferase structures in the calicheamicin biosynthetic pathway reveals the origin of regioselectivity. *Proc Natl Acad Sci U S A.* 2011; 108(43):17649–17654. [PubMed: 21987796]
- Court MH. Isoform-selective probe substrates for in vitro studies of human UDP-glucuronosyltransferases. *Methods Enzymol.* 2005; 400:104–116. [PubMed: 16399346]
- Coutinho PM, Deleury E, Davies GJ, Henrissat B. An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol.* 2003; 328(2):307–317. [PubMed: 12691742]
- Dong D, Wu B. Substrate Selectivity of Human Cytochrome P450s Predicted from Crystal Structures and in silico Modeling. *Drug Metab Rev.* 2012; 44(1):1–17. [PubMed: 22242930]
- Ekins S, de Groot MJ, Jones JP. Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome p450 active sites. *Drug Metab Dispos.* 2001; 29(7):936–944. [PubMed: 11408357]

- Ethell BT, Ekins S, Wang J, Burchell B. Quantitative structure activity relationships for the glucuronidation of simple phenols by expressed human UGT1A6 and UGT1A9. *Drug Metab Dispos.* 2002; 30(6):734–738. [PubMed: 12019203]
- He XZ, Li WS, Blount JW, Dixon RA. Regioselective synthesis of plant (iso)flavone glycosides in *Escherichia coli*. *Appl Microbiol Biotechnol.* 2008; 80(2):253–260. [PubMed: 18568307]
- He XZ, Wang X, Dixon RA. Mutational analysis of the Medicago glycosyltransferase UGT71G1 reveals residues that control regioselectivity for (iso)flavonoid glycosylation. *J Biol Chem.* 2006; 281(45):34441–34447. [PubMed: 16982612]
- Hu Y, Walker S. Remarkable structural similarities between diverse glycosyltransferases. *Chem Biol.* 2002; 9(12):1287–1296. [PubMed: 12498881]
- Kaivosaari S, Finel M, Koskinen M. N-glucuronidation of drugs and other xenobiotics by human and animal UDPglucuronosyltransferases. *Xenobiotica.* 2011; 41:652–669. [PubMed: 21434773]
- Kerdpin O, Mackenzie PI, Bowalgaha K, Finel M, Miners JO. Influence of N-terminal domain histidine and proline residues on the substrate selectivities of human UDP-glucuronosyltransferase 1A1, 1A6, 1A9, 2B7, and 2B10. *Drug Metab Dispos.* 2009; 37(9):1948–1955. [PubMed: 19487247]
- Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem.* 2008; 77:521–555. [PubMed: 18518825]
- Lewinsky RH, Smith PA, Mackenzie PI. Glucuronidation of bioflavonoids by human UGT1A10: structure-function relationships. *Xenobiotica.* 2005; 35(2):117–129. [PubMed: 16019943]
- Li D, Fournel-Gigleux S, Barré L, Mulliert G, Netter P, Magdalou J, Ouzzine M. Identification of aspartic acid and histidine residues mediating the reaction mechanism and the substrate specificity of the human UDP-glucuronosyltransferases 1A. *J Biol Chem.* 2007b; 282(50):36514–36524. [PubMed: 17956868]
- Li L, Modolo LV, Escamilla-Trevino LL, Achnine L, Dixon RA, Wang X. Crystal structure of Medicago truncatula UGT85H2--insights into the structural basis of a multifunctional (iso)flavonoid glycosyltransferase. *J Mol Biol.* Jul. 2007a; 370(5):951–963.
- Mackenzie PI, Bock KW, Burchell B, Guillemette C, Ikushiro S, Iyanagi T, Miners JO, Owens IS, Nebert DW. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet Genomics.* 2005; 15(10):677–685. [PubMed: 16141793]
- Miley MJ, Zielinska AK, Keenan JE, Bratton SM, Radomska-Pandya A, Redinbo MR. Crystal structure of the cofactor-binding domain of the human phase II drug-metabolism enzyme UDP-glucuronosyltransferase 2B7. *J Mol Biol.* 2007; 369(2):498–511. [PubMed: 17442341]
- Mittler M, Bechthold A, Schulz GE. Structure and action of the C-C bond-forming glycosyltransferase UrdGT2 involved in the biosynthesis of the antibiotic urdamycin. *J Mol Biol.* 2007; 372(1):67–76. [PubMed: 17640665]
- Modolo LV, Blount JW, Achnine L, Naoumkina MA, Wang X, Dixon RA. A functional genomics approach to (iso)flavonoid glycosylation in the model legume Medicago truncatula. *Plant Mol Biol.* 2007; 64(5):499–518. [PubMed: 17437063]
- Modolo LV, Li L, Pan H, Blount JW, Dixon RA, Wang X. Crystal structures of glycosyltransferase UGT78G1 reveal the molecular basis for glycosylation and deglycosylation of (iso)flavonoids. *J Mol Biol.* 2009; 392(5):1292–1302. [PubMed: 19683002]
- Mulichak AM, Losey HC, Lu W, Wawrzak Z, Walsh CT, Garavito RM. Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc Natl Acad Sci U S A.* 2003; 100(16):9238–9243. [PubMed: 12874381]
- Mulichak AM, Losey HC, Walsh CT, Garavito RM. Structure of the UDP-glycosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure.* 2001; 9(7):547–557. [PubMed: 11470430]
- Mulichak AM, Lu W, Losey HC, Walsh CT, Garavito RM. Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. *Biochemistry.* 2004; 43(18):5170–5180. [PubMed: 15122882]
- Noguchi A, Saito A, Homma Y, Nakao M, Sasaki N, Nishino T, Takahashi S, Nakayama T. A UDP-glucose:isoflavone 7-O-glucosyltransferase from the roots of soybean (*glycine max*) seedlings.

- Purification, gene cloning, phylogenetics, and an implication for an alternative strategy of enzyme catalysis. *J Biol Chem.* 2007; 282(32):23581–23590. [PubMed: 17565994]
- O'Dwyer PJ, Catalano RB. Uridine diphosphate glucuronosyltransferase (UGT) 1A1 and irinotecan: practical pharmacogenomics arrives in cancer therapy. *J. Clin. Oncol.* 2006; 24:4534–4538. [PubMed: 17008691]
- Offen W, Martinez-Fleites C, Yang M, Kiat-Lim E, Davis BG, Tarling CA, Ford CM, Bowles DJ, Davies GJ. Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.* 2006; 25(6):1396–1405. [PubMed: 16482224]
- Osmani SA, Bak S, Møller BL. Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling. *Phytochemistry.* 2009; 70(3):325–347. [PubMed: 19217634]
- Patana AS, Kurkela M, Finel M, Goldman A. Mutation analysis in UGT1A9 suggests a relationship between substrate and catalytic residues in UDP-glucuronosyltransferases. *Protein Eng Des Sel.* 2008; 21(9):537–543. [PubMed: 18502788]
- Radomska-Pandya A, Bratton SM, Redinbo MR, Miley MJ. The crystal structure of human UDP-glucuronosyltransferase 2B7 C-terminal end is the first mammalian UGT target to be revealed: the significance for human UGTs from both the 1A and 2B families. *Drug Metab Rev.* 2010; 42(1): 133–144. [PubMed: 19821783]
- Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X. Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *Plant Cell.* 2005; 17(11):3141–3154. [PubMed: 16214900]
- Smith PA, Sorich MJ, Low LS, McKinnon RA, Miners JO. Towards integrated ADME prediction: past, present and future directions for modelling metabolism by UDP-glucuronosyltransferases. *J Mol Graph Model.* 2004; 22(6):507–517. [PubMed: 15182810]
- Smith PA, Sorich MJ, McKinnon RA, Miners JO. Pharmacophore and quantitative structure-activity relationship modeling: complementary approaches for the rationalization and prediction of UDP-glucuronosyltransferase 1A4 substrate selectivity. *J Med Chem.* 2003a; 46(9):1617–1626. [PubMed: 12699380]
- Smith PA, Sorich MJ, McKinnon RA, Miners JO. In silico insights: chemical and structural characteristics associated with uridine diphosphate-glucuronosyltransferase substrate selectivity. *Clin Exp Pharmacol Physiol.* 2003b; 30(11):836–840. [PubMed: 14678246]
- Sorich MJ, Miners JO, McKinnon RA, Smith PA. Multiple pharmacophores for the investigation of human UDP-glucuronosyltransferase isoform substrate selectivity. *Mol Pharmacol.* 2004; 65(2): 301–308. [PubMed: 14742671]
- Sorich MJ, Smith PA, McKinnon RA, Miners JO. Pharmacophore and quantitative structure activity relationship modelling of UDP-glucuronosyltransferase 1A1 (UGT1A1) substrates. *Pharmacogenetics.* 2002; 12(8):635–645. [PubMed: 12439224]
- Sorich MJ, Smith PA, Miners JO, Mackenzie PI, McKinnon RA. Recent advances in the in silico modelling of UDP glucuronosyltransferase substrates. *Curr Drug Metab.* 2008; 9(1):60–69. [PubMed: 18220572]
- Vashishtha SC, Hawes EM, McKay G, McCann DJ. Quaternary ammonium-linked glucuronidation of 1-substituted imidazoles: studies of human UDP-glucuronosyltransferases involved and substrate specificities. *Drug Metab Dispos.* 2001; 29(10):1290–1295. [PubMed: 11560872]
- Wu B, Basu S, Wang X, Hu M. Regioselective Sulfation and Glucuronidation of Phenolics: Insights into the Structural Basis. *Curr Drug Metab.* (E-pub). 2011d
- Wu B, Kulkarni K, Basu S, Zhang S, Hu M. First-pass metabolism via UDP-glucuronosyltransferase: a barrier to oral bioavailability of phenolics. *J Pharm Sci.* 2011a; 100(9):3655–3681. [PubMed: 21484808]
- Wu B, Morrow JK, Singh R, Zhang S, Hu M. Three-dimensional quantitative structure-activity relationship studies on UGT1A9-mediated 3-O-glucuronidation of natural flavonols using a pharmacophore-based comparative molecular field analysis model. *J Pharmacol Exp Ther.* 2011c; 336(2):403–413. [PubMed: 21068207]

- Wu B, Wang X, Zhang S, Hu M. Accurate Prediction of Glucuronidation of Structurally Diverse Phenolics by Human UGT1A9 Using Combined Experimental and In Silico Approaches. *Pharm Res.* (Accepted). 2012
- Wu B, Xu B, Hu M. Regioselective glucuronidation of flavonols by six human UGT1A isoforms. *Pharm Res.* 2011b Aug; 28(8):1905–1918. [PubMed: 21472492]
- Xie S, Chen Y, Chen S, Zeng S. Structure-metabolism relationships for the glucuronidation of flavonoids by UGT1A3 and UGT1A9. *J Pharm Pharmacol.* 2011; 63(2):297–304. [PubMed: 21235595]
- Yang M, Brazier M, Edwards R, Davis BG. High-throughput mass-spectrometry monitoring for multisubstrate enzymes: determining the kinetic parameters and catalytic activities of glycosyltransferases. *Chembiochem.* 2005; 6(2):346–357. [PubMed: 15678424]
- Zhang C, Bitto E, Goff RD, Singh S, Bingman CA, Griffith BR, Albermann C, Phillips GN Jr, Thorson JS. Biochemical and structural insights of the early glycosylation steps in calicheamicin biosynthesis. *Chem Biol.* 2008; 15(8):842–853. [PubMed: 18721755]
- Zhang Z, Kochhar S, Grigorov M. Exploring the sequence-structure protein landscape in the glycosyltransferase family. *Protein Sci.* 2003; 12(10):2291–2302. [PubMed: 14500887]

Figure 1. Panel A: Cartoon representation of glycosyltransferase GT-A and GT-B folds
 N-terminal domain is colored cyan and C-terminal domain orange. The GT-A fold consists of two dissimilar $\beta/\alpha/\beta$ domains, with the N-terminal domain binding the nucleotide sugar and the highly variable C-terminal domain binding the acceptor substrate. The GT-B fold consists of two similar Rossmann fold domains, with the highly variable N-terminal domain binding the acceptor substrate and the C-terminal domain binding the nucleotide sugar. The enzyme for GT-A fold demonstration is the nucleotide-diphospho-sugar transferase SpsA from *Bacillus subtilis* (PDB code: 1H7L). The enzyme for GT-B fold demonstration is the

triterpene UDP-glucosyltransferase UGT71G1 from *Medicago truncatula* (PDB code: 2ACW). **Panel B: Two main catalytic mechanisms of glycosyltransferase-mediated reaction: retention and inversion of the anomeric configuration.** Human UGTs are inverting glycosyltransferases, like all other members of the GT1 family. X = -OH, -COOH, -NH, -NH₂, or -SH.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 2. The different catalytic mechanisms for O- and N-glucuronidation in human UGTs
Panel A: Proposed catalytic mechanism for O-glucuronidation (so-called “serine hydrolase like mechanism”) involving two catalytic residues histidine and aspartate. **Panel B:** N-glucuronidation does not required proton abstraction because N-nucleophiles can readily develop positive charged amine as the transition state. The question mark denotes that the role of the “catalytic” dyad histidine-aspartate in N-glucuronidation remains unknown.

Human UGTs	UGT1A1	33	IPVDGSHWLSMLG ... FDVMLTDPFPLPCS	157
	UGT1A3	34	VPIDGSHWLSMRE ... FDVVLTDPVNLCA	158
	UGT1A4 *	34	VPTDGSPWLSMRE ... FDVVLTDPVNLCA	158
	UGT1A5	34	VPTDGSHWLSMRE ... FDVVLTDPFHLCA	158
	UGT1A6	32	VPQDGSHWLSMKD ... FDALFTDPALPCG	156
	UGT1A7	31	VPMDGSHWFTMQS ... FDAVFLDPFDACG	154
	UGT1A8	31	VPMDGSHWFTMQS ... FDAVFLDPFDACG	154
	UGT1A9	31	VPMDGSHWFTMRS ... FDAVFLDPFDNCG	154
	UGT1A10	31	VPMDGSHWFTMQS ... FDAVFLDPFDTCG	154
	UGT2B4	29	WPTEFSHWMNKT ... FDVVLADAVFPFG	157
Plant UGTs	UGT2B7	29	WAAEYSHWMNKT ... FDVIFADAIFFPCS	157
	UGT2B10 *	28	WAAEYSLWMNMKT ... FDIVFADAYLPCG	156
	UGT2B11	29	WAAEYSHWMNMKT ... FDIVFADAVFPFG	157
	UGT2B15	29	WPTEYSHWINMKT ... FDVILADALNPCG	158
	UGT2B17	29	WPTEYSHWINMKT ... FDVLLADAVNPCG	158
	UGT2B28	29	WTGEYSHWMNMKT ... FDIIFADAFFPCG	157
	UGT71G1	16	PAPGIGHLASALE ... VVGLVLDFFCVSM	127
Bacterial UGTs	UGT72B1	13	PSPGMGHLIPLVE ... PTALVVDLFGTDA	123
	UGT82H5	15	PYPVQGHINPLFK ... VTCLVSDCCMSFT	131
	UGT78G1	20	AFPPFGTHAAPLLS ... ITCLVTDAFFWFG	130
	VvGT1	14	AFPFSTHAAPLLA ... VSCLVADAFIWFA	125
	Gtfa *	7	GCGSRGDTEPLVA ... CDAVVTTGLLPAA	105
	GtfB *	7	TCGSRGDTEPLVA ... CAAVVTTGLLAAA	106
	GtfD *	7	VCGTRGDVEIGVA ... CAAVVAVGDLAAA	107
	CalG1	7	SLGTHGHTYPLLP ... PDLVVQEISN-YG	122
	CalG2 *	8	NVASHGLILPTLT ... PDLVLYDDFPFIA	112
	CalG3	5	SSPGIGHLFPLIQ ... PDLVVYEQGA-TV	120
Bacterial UGTs	CalG4	10	NVQGHGHVYPSLG ... PDLVVYDVFPFIA	114
	OleD	13	SIAAHGHVNPSLE ... PDLVLHDITSYPA	116
	OleI	19	NIPGHGHVNPSLG ... PDLIVYDIASWPA	122
	UrdGT2 *	7	AAGSPATVFALAP ... PDLIVGGTMSYVA	120

Figure 3. A sequence alignment between human UGTs and other GT1 enzymes with solved crystal structures

The putative catalytic base is highlighted in red box and the corresponding acid in blue box.

Stars denote the enzymes which do not use histidine as the catalytic base.

Figure 4. The common features pharmacophore for human UGT1A1, 1A4, and 1A9 (Miners et al., 2004)

Panel A: The common features pharmacophore include two hydrophobic regions (H) in addition to the glucuronidation site (G). **Panel B:** the pharmacophore was mapped to selective substrates for UGT1A1, 1A4 and 1A9.

Figure 5. A homology model for human UGT1A9 and the superimposition of the CoMFA/CoMSIA steric maps with the modeled active site (Wu et al., 2012)

Panel A shows the constructed model for UGT1A9. The model consists of N- (in gray) and C-terminal (in light green) domains. The N- and C-terminal domains contain central stranded parallel sheets flanked by α -helices on both sides. The substrate binding pocket is almost entirely formed by the N-terminal residues, although some C-terminal residues also contributed to the formation of the pocket. The pocket is primarily formed by LoopN1, Na1, Na3-2, LoopN4, Na5-1, Na5-2, Loop C1 and Loop C5. The cofactor is present at the left

side of pocket. The catalytic residue histidine (in green stick model) was located at the start of helix N α 1. The substrate kaempferol (in a 3-OH catalysis mode) is shown in stick-and-ball model. **Panel B/C:** Superposition of the CoMFA (**B**) and CoMSIA (**C**) steric maps over the active site of the homology-modeled UGT1A9 structure based on a simulated binding model of kaempferol (3-OH). The UGT1A9 protein is shown in a stick model. Kaempferol is indicated in a ball-and-stick model and the cofactor is shown in a ball-and-stick model with a molecular surface. Green: Areas in which bulky groups are sterically favorable for glucuronidation; Yellow: Areas in which bulky groups are unfavorable for glucuronidation.

Figure 6. Comparisons of the active sites (in a green solid surface) of crystal structures for 5 plant UGTs: (A) UGT71G1 (PDB code, 2ACW), (B) UGT72B1 (PDB code, 2VCE), (C) UGT78G1 (PDB code, 3HBF), (D) UGT85H2 (PDB code, 2PQ6), and (E) VvGT1 (PDB code, 2C1Z)

The volume of UGT72B1 active site is estimated at $41 \text{ \AA}^3$, much smaller than those of UGT71G1, UGT78G1, UGT85H2, or VvGT1. Please note that estimates of the active site volumes (indicated by question marks) for the latter are inaccurate (greatly less than the actual values) as the active sites are open to the solvent. The bars denote a rough distance between the active site and the helix Na5, suggesting that residues from Na5 creates a more significant steric hindrance for UGT71G1 than for UGT78G1, UGT85H2, or VvGT1.

Figure 7. Comparisons of the active sites of GT1 enzymes reveals significant conformational differences are required to accommodate distinct substrates

Panel A: An active site comparison of UGT72B1-TCP (in cyan) and VvGT1-kaempferol (in white) complexes. **Panel B:** An active site comparison of OleI-oleandomycin (in blue) and VvGT1-kaempferol (in white) complexes. The PDB codes for the crystal structures can be found in Table 1. Star denotes the unique loop C2 in UGT72B1. Dashed line denotes the hydrogen bond. TCP: 2,4,5-trichlorophenol.

Figure 8. Accommodation of substrates in GT1 enzymes requires conformational changes in the protein in relation to the substrate free enzymes

The changes mainly occur at the side chains of residues forming the active site. **Panel A:** a comparison of UGT72B1 binding site for TCP binding (in cyan) with the substrate-free binding site (in white). **Panel B:** a comparison of UGT78G1 structures for myricetin binding (in purple) with the substrate-free binding site (in white). **Panel C:** A comparison of VvGT1 structures for kaempferol binding (in cyan) with the substrate-free binding site (in white).

Panel D: A comparison of VvGT1 structures for quercetin binding (in purple) with the

substrate-free binding site (in white). The PDB codes for the crystal structures can be found in Table 1. Star denotes the site of glycosylation. Dashed line denotes the hydrogen bond. TCP: 2,4,5-trichlorophenol.

Human UGT1A1: 3'-OH > 7-OH > 3-OH

Figure 9. A hypothetical binding model of flavonols (quercetin is shown as an example) to human UGT1A1

Panels A-B show the molecular basis for regioselectivity of 4 plant UGTs towards quercetin deduced from crystal structures and mutagenesis studies (Osmani et al., 2009). In the presence of steric hindrance from Na5 (i.e., F148), positioning of 3-OH of quercetin in UGT71B1 is disfavored. In the absence of significant steric hindrance from Na5, positioning of 3-OH of quercetin in UGT78G1, UGT85H2, and VvGT1 is favored. Please also refer to Figure 6 for the comparisons of substrate-binding pockets among plant UGTs. Panel B (VvGT crystal structure is used) is shown for a better understanding of the geometry of the substrate-binding pocket (i.e., the green meshed surface) within the plant UGT protein, as well as its relative coordinates to the helix Na5 (situates at the end of the pocket), catalytic residue histidine, cofactor, and solvent. **Panel C** shows the hypothetical binding

modes of quercetin to human UGT1A1. The inner part (or the end) of substrate-binding pocket provides steric bulks (possible aromatic rings), which disfavor the positioning of flavonols for 3-OH catalysis. Comparing to 7-OH, 3'-OH is more preferred by the potential H-bond formation (dashed green line) between 4'-OH and the active site residues.

Table 1

List of available GT1 crystal structures, the structures can be found at <http://www.rcsb.org/pdb/home/home.do>.

Glycosyltransferase	Crystal name	In complex with	Reference
UGT71G1 (GT29H)	2ACV	UDP	Shao et al., 2005
	2ACW	UDP-glucose	Shao et al., 2005
UGT82H5 (GT67A)	2PQ6		Li et al., 2007
UGT78G1 (GT83F)	3HBJ	UDP	Modolo et al., 2009
	3HBF	UDP, Myricetin	Modolo et al., 2009
VvGT1	2CIX	UDP	Offen et al., 2006
	2CIZ	UDP-2-fluoro-glucose, kaempferol	Offen et al., 2006
	2C9Z	UDP, quercetin	Offen et al., 2006
UGT72B1	2VCE	UDP-2-fluoro glucose, 2,4,5-trichlorophenol	Brazier-Hicks et al., 2007
	2VCH	UDP	Brazier-Hicks et al., 2007
	2VG8	UDP, Tris buffer	Brazier-Hicks et al., 2007
GtfD	1RRV	TDP	Mulichak et al., 2004
GtfB	1IIR		Mulichak et al., 2001
GtfA	1PN3	TDP, Desvancosaminyl vancomycin	Mulichak et al., 2003
	1PNV	TDP	Mulichak et al., 2003
CalG1	3OTG	TDP	Chang et al., 2011
	3OTH	TDP, Calicheamicin α_3^I	Chang et al., 2011
CalG2	3IAA	TDP	Chang et al., 2011
	3RSC	TDP, Calicheamicin T_0	Chang et al., 2011
CalG3	3D0Q	3[N-morpholino]propane sulfonic acid	Zhang et al., 2008
	3D0R	PEG4000	Zhang et al., 2008
	3OTI	TDP, Calicheamicin T_0	Zhang et al., 2008
CalG4	3IA7		Chang et al., 2011
OleD	2IYF	UDP, Erythromycin	Bolam et al., 2007
OleI	2IYA	UDP, Oleandomycin	Bolam et al., 2007
UrdGT2	2P6P		Mittler et al., 2007

Table 2

List of the substrates of plant glucosyltransferases

Plant UGTs	Substrates*	References
UGT71G1	Kaempferol; Quercetin (3'-OH); Genistein (7-OH); biochanin A; hederagenin medicagenic acid; Isoliquiritigenin, Coumestrol, Liquiritigenin, Naringenin, Apigenin, Luteolin, Tricin, , Daidzein, Formononetin, Irisolidone, Medicarpin	Shao et al., 2005; He et al., 2006, 2008; Modolo et al., 2007
UGT82H5	Isoliquiritigenin, Kaempferol; Quercetin, Myricetin, Biochanin A	Modolo et al., 2007
UGT78G1	Pelargonidin (3-OH), Isoliquiritigenin, Coumestrol, Naringenin, Apigenin (7-OH), Chrysoeriol, Luteolin, Tricin, Kaempferol (3-OH); Quercetin (3-OH), Myricetin, Afmosin, Biochanin A (7-OH), Daidzein (7-OH), Formononetin (7-OH), Genistein (7-OH), Irisolidone, Medicarpin, Cyanidin (7-OH)	Modolo et al., 2009
VvGT1	Cyanidin (3-OH), Kaempferol (3-OH), Quercetin (3-OH), Umbelliferone, 4-methylumbelliferone, Scopoletin, Esculetin, Baicalein, Luteolin, Fisetin, Genistein, Taxifolin, Cyanidin chloride, Caffeic acid	Offen et al., 2006
UGT72B1	Umbelliferone; 4-methylumbelliferone, 3,4-dichloroaniline; 3,4,5-Trichloroaniline; 3,4-dihydroxyl benzoic acid; 2,4,5-trichlorophenol; 2,5-dihydroxyl benzoic acid; Triclosan Scopoletin, Esculetin, Dihydrojasmonic acid	Yang et al., 2005; Brazier-Hicks and Edwards, 2005; Brazier-Hicks et al., 2007a,b

* Content in the parenthesis indicates the major site of glycosylation