IDENTIFICATION OF RESIDUES THAT CONFER SUGAR SELECTIVITY TO UDP GLYCOSYLTRANSFERASE 3A (UGT3A) ENZYMES.

Robyn Meech, Anne Rogers, Lizhe Zhuang, Benjamin C. Lewis, John O. Miners and Peter I. Mackenzie

From Department of Clinical Pharmacology, Flinders University School of Medicine, Flinders Medical Centre, Bedford Park, SA 5042, Australia.

Running Head: Sugar specificity of UGTs

Address correspondence to Robyn Meech, Clinical Pharmacology, Flinders Medical Centre, Bedford Park, SA 5042, Australia. Fax: +61-8-82045114; Email: Robyn.Meech@flinders.edu.au

Background: Conjugation of sugars to chemicals by UGTs is a critical detoxification mechanism.
Result: A single amino acid defines the differential sugar specificities of two related UGTs.
Conclusion: Change of a single amino acid during primate evolution has generated a new capacity for small molecule glycosidation.
Significance: Determinants of UGT sugar selectivity are currently poorly understood; novel glycosidation pathways may have important metabolic roles.

SUMMARY

Recent studies in this laboratory characterized the UGT3A family enzymes, UGT3A1 and UGT3A2, and showed that neither uses the traditional UGT co-substrate UDP-glucuronic acid. Rather, UGT3A1 uses N-acetylglucosamine as preferred sugar donor and UGT3A2 uses UDP-glucose. The enzymatic characterization of UGT3A mutants, structural modelling, and multispecies gene analysis have now been employed to identify a residue within the active site of these enzymes that confer their unique sugar preferences. An asparagine (N391) residue in the UGT signature sequence of UGT3A1 is necessary for utilization of UDP-N-acetylglucosamine. Conversely, a phenylalanine (F391) residue in UGT3A2 favors UDP-glucose use. Mutation of N391 to F in UGT3A1 enhances its ability to utilize UDP-glucose and completely inhibits its ability to use UDP-N-acetylglucosamine. An analysis of homology models docked with UDP-sugar donors indicates that N391 in UGT3A1 is able to accommodate the N-acetyl group on C2 of UDP N-acetylglucosamine so that the anomeric carbon atom (C1) is optimally situated for catalysis involving H35. Replacement of N by F at 391 disrupts this catalytically-productive orientation of UDP N-acetylglucosamine but allows a more optimal alignment of UDP-glucose for sugar donation. Multispecies sequence analysis reveals that only primates possess UGT3A sequences containing the N391 residue, suggesting that other mammals may not have the capacity to N-acetylglucosaminidate small molecules. In support of this hypothesis, N391-containing UGT3A forms from two non-human primates were found to use UDP-N-acetylglucosamine, while UGT3A isoforms from non-primates could not use this sugar donor. This work gives new insight into the residues that confer sugar specificity to UGT family members and suggests a primate-specific innovation in glycosidation of small molecules.

Glycosidation of lipophilic chemicals by the UDP glycosyltransferase (UGT) superfamily is an important detoxification pathway in all vertebrates (1). Glycosidation increases the water-solubility of the acceptor substrate, facilitating its excretion and/or otherwise altering its biological reactivity (2-5). Acceptors are structurally diverse and include steroid hormones, bile acids, biogenic amines, plant and bacterial metabolites, carcinogens and many therapeutic drugs (6). Thus glycosidation protects cells against exogenous toxins and accumulations of potentially toxic by-products

of metabolism. It can also play a role in modulating signalling pathways such as those mediated by steroid hormones. The glycosyl donor (co-substrate) is usually a UDP-hexose, typically UDP-glucuronic acid (UDP-GlcUA), UDP-glucose (UDP-Glc), UDP-xylose (UDP-Xyl), UDP-galactose, (UDP-Gal) or UDP N-acetylglucosamine (UDP-GlcNAc). During the conjugation reaction, the UDP-α-bond between UDP and the hexose moiety is converted into a β-bond between the acceptor and the sugar via a S_N2 mechanism, to form a β-D-glycoside.

In humans, four UGT families have been identified; UGT1, UGT2 (divided into sub-families, 2A and 2B), UGT3 and UGT8 (6). The UGT1 enzymes are encoded by a single genomic locus on chromosome 2q37, which contains multiple enzyme specific exons 1A and a shared set of exons 2-5 (7). Differential promoter usage and splicing produces mRNAs for 9 separate UGT1A enzymes, each having a unique N-terminal domain encoded by an exon 1A and an identical C-terminal domain encoded by exons 2-5. The UGT2 family is divided into the UGT2A and UGT2B sub-families. UGT2A1 and 2A2 are encoded in a similar fashion to the UGT1A family, and have identical C-terminal domains encoded by a shared set of 5 exons. The remaining members of the UGT2 family are encoded by separate genes of 6 exons each, that are arrayed along chromosome 4q13 (6). The UGT3A family has two members denoted *UGT3A1* (8) and *UGT3A2* (9) that are arranged as a direct repeat on chromosome 5p13. The human UGT8 family contains only one gene *UGT8A1* located on chromosome 4q26.

The UGT1 and UGT2 enzymes use UDP-GlcUA as their preferred sugar donor, although there are examples where other UDP-sugars are used. For example, UGT2B7 and UGT1A1 can utilize UDP-Glc, and UGT1A1 can use UDP-Xyl (10-12). However, the activities of UGT1 and UGT2 family enzymes with these alternate UDP-sugars are much lower than those with UDP-GlcUA, and are typically restricted to specific aglycone substrates. There is no evidence that UGT1 and UGT2 forms can use UDP-Gal or UDP-GlcNAc. UGT8 catalyses the transfer of galactose from UDP-Gal to ceramide in the production of glycosphingolipids; no other sugar donors or acceptors have been identified for this enzyme. Currently it is believed that UGT8 has an exclusively biosynthetic role and is not involved in detoxification (13).

We recently characterized the substrate specificities of the UGT3A enzymes. In striking contrast to the UGT1 and UGT2 families, we found that neither UGT3A enzyme can utilize the prototypic UGT co-substrate UDP-GlcUA. UGT3A1 uses UDP-GlcNAc to conjugate several substrates including bile acids, steroids and bioflavones (8). This sugar selectivity is unique among vertebrate UGTs. UGT3A2 preferentially utilizes UDP-Glc with a range of substrates including bioflavones and estrogens (9). The significance of these unusual sugar specificities remains unclear, although it is possible that the N-acetylglucosamine and glucose conjugates have important biological roles in vivo. UGT3A2, in particular, is expressed at very low levels in human liver, kidney and gastrointestinal tract suggesting that its major roles may be not in elimination of exogenous chemicals, but in endogenous metabolism.

UGTs can be divided into two domains and typically show much greater sequence conservation within their C-terminal halves than their N-terminal halves. This is consistent with recognition of diverse aglycone substrates by the N-terminal domain and binding of the UDP-sugar donor by the C-terminal domain. While sequence determinants of aglycone specificity have been identified by domain swapping and mutagenesis experiments, there has been limited progress in identifying residues involved in sugar specificity, in part because all UGT1 and 2 enzymes preferentially use UDP-GlcUA and have only minor activities with other sugars. The recent discovery of differential sugar usage in the UGT3A family now provides a unique opportunity to address this issue. Using cross-species sequence comparisons and site directed mutagenesis, we have identified a key residue in the putative sugar-binding domain that confers differential sugar specificities to the UGT3A proteins. We also provide evidence that the use of UDP-GlcNAc as a sugar donor is unique to primate UGT3A proteins.

EXPERIMENTAL PROCEDURES

Materials- Radioactive and non-radioactive UDP-sugars were obtained from the following: [C-14]UDP-glucose and -N-acetylglucosamine (GE Healthcare), UDP-glucose, -N-acetylglucosamine (Sigma-Aldrich). All other reagents and solvents were of analytical reagent grade.

Production of UGT3A1 and UGT3A2 mutants - Site-directed mutagenesis was performed using human UGT3A1 and UGT3A2 cDNAs encoded in the PCR2.1 vector and the Quick-Change mutagenesis method. N391 in UGT3A1 was mutated to F391 using the following primers: forward 5'-CCCATGGTGGGATTACCAGTCTTTGGAGACCAGCATGGAAACATGG-3' and reverse 5'-CCATGTTTCCATGCTGGTCTCCAAAGACTGGTAATCCCACCATGGG-3'. F391 in UGT3A2 was mutated to N391 using the following primers: forward 5'-CCCATGGTGGGGATCCCTCTCAATGGAGACCAGCCTGAAAACATGG-3' and reverse 5'-CCATGTTTTCAGGCTGGTCTCCATTGAGAGGGATCCCCACCATGGG-3'. The mutated products were verified by re-sequencing of the complete coding region and then subcloned into the pEF-IRESpuro6 expression vector which contains a puromycin resistance gene (14). Expression vectors were transfected into human embryonic kidney (HEK293T) cells and mixed cell populations stably expressing the mutant proteins were selected with puromycin (2 μg/ml). Lysates were prepared for assays in buffer containing 10 mM Tris, 1 mM EDTA pH 7.6.

Western blotting- Proteins in lysates from HEK293T cells stably expressing wildtype and mutant UGT3A cDNAs were separated on SDS-polyacrylamide gels and transferred to nitrocellulose membranes as previously described (8,15). UGT3A wildtype and mutated proteins were detected with UGT3A1 and UGT3A2 antibodies generated previously (8,9) and a secondary goat anti-rabbit antibody conjugated with peroxidase (Neomarkers). Immunocomplexes were visualized with the Supersignal West Pico chemiluminescent kit (ThermoFisher Scientific) and quantified with the LAS4000 scanner (GE Healthcare).

Enzyme assays- For assays to assess co-substrate preference, glycosidation reactions were performed as previously described (8). Briefly, incubations at 37^{o}C for 1 hr, contained 100 mM phosphate buffer, pH 7.5, 4 mM magnesium chloride, 100 μg HEK293T cell lysate, 200 μM aglycone and 2 mM [C-14]UDP-sugar (0.1 μCi/mMole). Radioactive products were separated by thin layer chromatography (8) and quantified by exposure to a Phosphor Screen (Molecular Dynamics), which was scanned with a Typhoon 9400 scanner (GE Healthcare). Standard curves with known amounts of [C-14]UDP-sugar were constructed to quantify product formation. For comparisons of activities between lysates expressing UGT3A proteins, activities relative to the amount of UGT protein present in the lysate, as determined by Western blotting, were used.

In silico modeling and docking- A human UGT3A1 homology model was constructed using FUGUE and ORCHESTRAR (SYBYL-X 1.3, Tripos™). Homologs identified by FUGUE were refined against profile Hidden Markov Model (16) data constructed from the UGT3A1 coding sequence. Crystal templates of grape UDP-glucose flavanoid 3-*O* glucosyltransferase (2C1X, 2C1Z, 2C9Z), glycosyltransferase UGT78G1 from Medicago truncatula (3HBF, 3HBJ), and the cofactor-binding domain of UGT2B7 (2O6L) (17) were obtained from the Brookhaven Protein Data Bank (PDB) and sequence alignments refined using Cn3D relative to secondary structure. The highest acceptable inter-Cα distance between equivalent residues within a sequence conserved region was 1.5Å, with a RMS difference of 0.00001Å considered significant. Loops not modeled were found by loop threading all crystal structures of the HOMSTRAD database (SYBYL-X 1.3, Tripos™). Refinement of the model was achieved by undertaking independent minimizations of biopolymer hydrogens, side-chains, the biopolymer omitting Cα carbons,

and finally the biopolymer as a whole. Each independent minimization allowed a maximum of 10,000 iterations using the method reported by Powell (18) with a termination gradient of <0.05 kcal/(mol)(Å). Automated docking of UDP-GlcNAc and UDP-Glc was achieved using the Surflex-Dock (SFXC) suite (SYBYL-X 1.3, Tripos™). Co-substrate docking was consensus scored (CScore) relative to protein interactions with hydrogen flexibility allowed. The relative strengths of co-substrate and protein covalent force fields were set to 1.00 and 0.10 respecively. Ring flexibility was additionally allowed when generating the molecular co-substrate fragments and minimized using a BGFS quasi-Newton method and an internal Dreiding force field (19). In silico mutants were generated by substiution of the desired amino acid. Substituted residues were energy minimized as a subset of the entire protein molecule using the Powell conjugate gradient method with an energy cutoff set to 0.05 kcal/mol.Å. A 'hot region' of 6Å surrounding the substituted residue was established where the sidechains of all residues were minimized. A further 'intermediate region' of 12Å was generated to set the minimization environment without side-chain movement. Minimization by this method allowed changes in the energetic forces experienced by residues that either adjoin or neighbor the substituted amino acid. Distance measurements were collected to characterize the orientation of docked co-substrate poses. Distances were measured between the anomeric carbon (C1) of the co-substrate sugar moiety (i.e. the site of S_N2 attack) and the N2 nitrogen of the catalytic base, H35.

Cloning and expression of Non-human UGT3A cDNAs – Mouse Ugt3a1 and Ugt3a2 cDNAs were obtained by reverse transcription of neonatal (up to one week old) mouse liver or kidney RNA. Briefly, first strand cDNA was synthesized with the Superscript™ First Strand Synthesis System (Invitrogen). The Ugt3a1 coding region was amplified from the cDNA using the forward primer, 5'-CGGAATTC*ATG*GCTGCACATCGGAGTTGGC-3' and the reverse primer, 5'-CGGAATTC*TTA*TGCCTGCTTGACCTTCCTTG-3'. The mouse Ugt3a2 coding region was amplified using the forward primer, 5'-CGGAATTC*ATG*GCAGCACATCGGCGTTGG-3' and the reverse primer, 5'-CGGAATTC*TTA*TGCCTCCTTGACCTTCGT-3'. The initiation and stop codons in the forward and reverse primers respectively, are in italics. Gene nomenclature follows that of the NCBI Gene database. PCR was performed in a volume of 50 µl with 200 ng cDNA, 0.5 µM of the forward and reverse primers and the DNA polymerase, Pfu Turbo (Stratagene). The cycling parameters consisted of one cycle at 95°C for 2 min, then 35 cycles of 95°C for 0.5 min, 60°C for 0.5 min, 72°C for 2 min followed by a single 5-min cycle at 72°C. PCR products were excised from a 1% agarose gel, purified using the QIAquick gel extraction kit (Qiagen), and cloned into the pCR2.1 shuttle vector (Invitrogen) for sequencing. Both cDNAs were subcloned into the pEF-IRESpuro6 expression vector.

Full length cDNAs for chimp (Pan troglodytes) UGT3A1 (XM_526949), rhesus (Macaca mulatta) UGT3A1 (XM_001093373) and UGT3A2 (XM_001093598), Cow (Bos taurus) UGT3A2 (NM_001076087), and rabbit (Oryctolagus cuniculus) UGT3A (XM_002714108) were obtained from Genescript in the pUC57 shuttle vector. The cDNAs were excised and subcloned into the pEF-IRESpuro6 expression vector. Sequencing confirmed the identities of each clone.

The human and chimp UGT3A2 proteins differ by only one amino acid (N309 in human is Y309 in chimp). To obtain a cDNA with the equivalent sequence to chimp UGT3A1, we performed site directed mutagenesis of the human UGT3A2 cDNA to introduce the chimp-specific amino acid (Y309). The mutagenesis primer sequences were as follows: forward 5-GTCAGTATCCGGAAATCTTCAAGGAG-3' and reverse 5'-CGGATACTGACAGGTGTTCACCATG-3'.

All non-human UGT cDNAs were transfected into HEK293T cells. Either transiently-expressing lysates, or lysates from puromycin-selected cell populations stably-expressing the UGT of interest, were used in assays.

Analysis of endogenous Ugt3a gene expression in mouse tissue – To measure the levels of endogenous Ugt3a1 and Ugt3a2 mRNAs in neonatal mouse livers and kidneys, we used quantitative RT-PCR with the following primers: Ugt3a1 forward 5'-ACCGTGTGTCGCAAATTCTG-3' and reverse 5'-ACCTGGTATGATGAGTTTTCC-3'; Ugt3a2 forward 5'-TTCTCATGAGCTTCCTTTTCC-3' and reverse 5'- GCGACACACGGCTTATCAC-3'. The raw data were converted to copy number using standard curves constructed from the cloned cDNAs. Data were then normalized to the mRNA abundance of the housekeeping gene ribosomal protein S26 (RPS26). Reactions used the Qiagen SYBR-green reagent or GoTaq (Promega) and the Rotor-Gene 300 (Corbett Life Sciences) thermal cycler.

Preparation of mouse liver and kidney microsomes – Livers and kidneys from 2-3 week old mice were snap frozen on dry ice. Microsomes were prepared according to (20) and assays were performed as described above.

RESULTS

Functional analysis of residue 391 in human UGT3A1 and UGT3A2. We have previously shown that UGT3A1 is unique among human UGTs in using UDP-GlcNAc as its preferred sugar donor (8), while human UGT3A2 utilizes UDP-Glc as its preferred co-substrate (9). The UGT signature sequence is a 30 amino acid region that is highly conserved in all UGTs and implicated in the binding of UDP-sugars (see (6) and references therein). When the signature sequences of all human UGTs are aligned, the residue 391 is clearly divergent; this residue is asparagine (N) in UGT3A1 and phenylalanine (F) in all other human UGTs including UGT3A2 (Figure 1). We hypothesized that N391 may be important for conferring the unique ability of human UGT3A1 to utilize UDP-GlcNAc.

Site-directed mutagenesis was used to substitute N391 in UGT3A1 with phenylalanine (N391F), and conversely to substitute F391 in UGT3A2 with asparagine (F391N). The stable expression of both mutated proteins in HEK293T cells was confirmed using western blotting (not shown). The activity of both mutated and wildtype proteins was examined using either UDP-Glc or UDP-GlcNAc and the aglycone substrate genistein. Genistein was selected because it was previously found to be a good substrate for both UGT3A1 and UGT3A2 with their preferred sugar donors (8,9,21).

Using sensitive assays for glycosidation, we observed that wildtype UGT3A1 in fact has a weak inherent capacity to conjugate glucose to genistein (see Supplemental Figure 1). Mutation of UGT3A1 N391 to F391 (N391F) dramatically enhanced glucosidation activity and simultaneously abolished activity with UDP-GlcNAc (Table 1). Mutation of UGT3A2 F391 to N391 (F391N) reduced the enzyme's activity with UDP-Glc by around 150-fold (Table 1); however, it did not confer activity with UDP-GlcNAc.

Structure modelling and UDP-sugar docking. To examine the structural basis underpinning the importance of residue 391, homology models of UGT3A1 and the UGT3A1-N391F mutant were generated and docked with UDP-GlcNAc or UDP-Glc (Figs 2 and 3). The docking of UDP-GlcNAc to UGT3A1 revealed direct side-chain H-bonding interactions from E377 to both hydroxyl hydrogens of the ribose, N373 to two oxygens of the diphosphate, H369 and Q394 to the 3' hydroxyl oxygen of the ribose, and D393 to the 4' hydroxyl hydrogen of the sugar moiety. Numerous main-chain (both nitrogen and oxygen) H-bonding interactions provided further stabilization of the UDP moiety, these include: L352 and Q354 to uracil, and Q372, N373, and S374 to the diphosphate. W351 was in a position to form π-π and π-σ orbital interactions with the uracil. The residues utilized in orientating UDP-GlcNAc for nucleophillic attack by the aglycone are located within the UGT signature sequence (Figure 2).

All UGTs appear to have a histidine (H35 in UGT3A1) which acts as a catalytic base to facilitate activation of the sugar-acceptor group on the aglycone by deprotonation for subsequent nucleophilic attack on the anomeric carbon (C1) of the sugar donor (22). When UDP-GlcNAc is docked into UGT3A1 (Figure 3A), the distance

between the catalytic histidine and C1 is 4.02Å, which increases to 6.21 Å when UDP-Glc is docked into the protein (Figure 3B). In contrast, when UDP-GlcNAc or UDP-Glc are docked into UGT3A1-F391 (Figures 3C and 3D) the distances are 6.18Å and 3.60Å, respectively. Hence, the UGT3A1 proteins that are most active with their respective sugar donors (i.e. UGT3A1 with UDP-GlcNAc and UGT3A1-F391 with UDP-Glc) have shorter distances between the catalytic histidine and C1, compared to the UGT3A1 protein/UDP-sugar combinations that have little activity (i.e. UGT3A1 with UDP-Glc, and UGT3A1-F391 with UDP-GlcNAc).

Cloning and functional analysis of mouse Ugt3a1 and Ugt3a2. We examined UGT3A sequences in the genomes of multiple species to determine when the N391 residue may have evolved. In all species examined, UGT3A genes were located in regions of shared synteny, including the flanking genes LMBR1 domain containing-2 (*LMBRD2*) and calcyphosine-like (*CAPSL*). These flanking genes were used as markers to guide our identification of UGT3A genes and pseudogenes in each genome. We found that the human, chimp, gorilla, baboon, macaque, marmoset and tarsier genomes each encode one UGT3A form with an N391 residue and one with an F391 residue (Figure 4). The orang-utan genome appears to encode only one UGT3A form and that bears the N391 residue. The genomes of the gibbon, bushbaby (otolemur) and mouse lemur, and of non-primates including mouse, rat, rabbit, dog, cow, horse, panda, elephant, and platypus contained one or more UGT3A genes that each encodes the F391 residue (or in some cases L391) and no genes that encode the N391 residue. In general, the signature sequences of primate UGT3A paralogues show less similarity to one another than those of the non-primate UGT3A paralogues. For example, the human UGT3A1 and UGT3A2 proteins are only 82% identical within their signature sequences while the mouse Ugt3a1 and Ugt3a2 are 95% identical within their signature sequences.

Given the effect of the UGT3A1-N391F mutation on sugar preference, we hypothesized that only genes that encode N391 may efficiently utilize UDP-GlcNAc as a sugar donor in conjugation reactions. To test this hypothesis we first cloned mouse Ugt3a1 and Ugt3a2 cDNAs and examined their sugar specificity in the heterologous HEK cell expression system.

As shown in Table 2, mouse Ugt3a1 was active with 4-methylumbelliferone (4-MU), and ursodeoxycholic acid (UDCA) using UDP-Glc as the sugar donor. However, it showed no activity toward either substrate with UDP-GlcNAc. Mouse Ugt3a2 was also able to conjugate 4-MU and UDCA with UDP-Glc, although to a lesser extent than Ugt3a1, and it was also inactive with UDP-GlcNAc.

The arrangement of the UGT3A genes and their immediate neighbours in the human genome is *LMBRD2* - *UGT3A2* - *UGT3A1* – *CAPSL*. However, in the public mouse genome databases, the name *Ugt3a1* has been assigned to the gene most proximal to *Lmbrd2* (e.g. *Lmbrd2* - *Ugt3a1* - *Ugt3a2* – *Capsl*) (see (21). The products of the mouse *Ugt3a1* and *Ugt3a2* genes show the same degree of sequence similarity to the human *UGT3A1* product. Based on genomic position, we suggest that mouse *Ugt3a2* is in fact likely to be the homologue of human *UGT3A1*. In support of this notion, it was found that mouse *Ugt3a1* shows a similar pattern of expression to human *UGT3A2* and that mouse *Ugt3a2* expression is more similar to that of human *UGT3A1*. Specifically, the expression of mouse *Ugt3a2* is vastly higher in liver and kidney than that of *Ugt3a1* (Figure 4A). We have previously shown that human *UGT3A1* mRNA is abundant in human liver and kidney while human *UGT3A2* mRNA is barely detectable in liver and very low in kidney (8,9).

Given the likelihood that mouse Ugt3a2 is the homologue of human UGT3A1; we sought additional aglycone substrates for this enzyme to confirm its apparent inability to use UDP-GlcNAc. Several bile salts and related chemicals were investigated. Hyodeoxycholic acid (HDCA) was identified as the aglycone with which mouse Ugt3a2 had highest glucosidation activity. However, even with this substrate, Ugt3a2 showed no activity with UDP-GlcNAc as sugar donor (Figure 5B). This result further

supports the idea that both mouse Ugt3a enzymes can utilize only UDP-Glc.

To assess whether mouse may have any capacity for N-acetylglucosaminidation of small molecules, we tested microsomal preparations of mouse liver and kidney for endogenous conjugation activity using UDP-Glc or UDP-GlcNAc and a variety of aglycones including mycophenolic acid (MCA), HDCA, hesperetin, and 4-MU. Only glucose conjugates were detected (Figure 5C), further supporting the idea that the mouse does not express an UDP-N-acetylglucosaminyltransferase.

Activities of chimp, rhesus, cow and rabbit UGT3A proteins

To further test our hypothesis that only genes that encode N391 may utilize UDP-GlcNAc as a sugar donor in conjugation reactions, we obtained an assortment of clones encoding UGT3A enzymes from various primate and non-primate species and expressed them in HEK293T cell culture. Activity was examined using a variety of aglycone substrates and both UDP-Glc and UDP-GlcNAc as sugar donor. All recombinant UGT3A enzymes were catalytically active. However, only those UGT3A enzymes that contain the N391 residue were able to use UDP-GlcNAc as sugar donor (Figure 6). Those with a phenylalanine in this position were only active with UDP-Glc as sugar donor.

DISCUSSION

UDP-GlcUA recognition by human UGTs has been investigated by several laboratories since early photoaffinity labelling experiments suggested a binding site between amino acids 299 and 466 (23). Several residues within the UGT signature sequence including H371 and E379 (numbering from UGT1A6) (24) and the DQ motif that marks the end of the UGT signature sequence (D393, Q394 in UGT1A10) (25) have been shown to play important or essential roles in UDP-GlcUA binding (25). In contrast, there has been little progress in understanding the structural basis of *differential* sugar recognition by UGTs, as most mammalian UGTs show only minor activities with sugars other than UDP-GlcUA. Characterization of the divergent sugar preferences of the UGT3A enzymes provided us with an opportunity to identify residues that may confer selectivity for different UDP-sugars.

By analysis of UGT3A family enzymes we have shown that N391 within the signature sequence of UGT3A1 is essential for its capacity to utilize UDP-GlcNAc as a sugar donor: substituting phenylalanine at this position blocks its ability use UDP-GlcNAc and instead promotes utilization of UDP-Glc. Docking of various sugars into a homology model of wildtype UGT3A1 suggests that the catalytic histidine is significantly closer to the anomeric carbon of UDP-GlcNAc than UDP-Glc, which likely explains its greater activity with UDP-GlcNAc than with UDP-Glc.

While the N391F mutation altered the sugar preference of UGT3A1, the level of activity of UGT3A1-N391F with UDP-Glc was lower than that of UGT3A2 with UDP-Glc, or of wildtype UGT3A1 with UDP-GlcNAc. This suggests that additional residues determine the efficiency with which the preferred sugar is used. We attempted to identify such residues by swapping signature sequences and adjacent segments between UGT3A1 and UGT3A2 (not shown). Some of these chimeric enzymes were poorly expressed suggesting that they may be structurally unstable. However, even among the variant UGT3A1 proteins that were efficiently expressed, we did not find any that had greater activity with UDP-Glc than the UGT3A1-N391F mutant (not shown). Thus the residue at position 391 is the major contributor to sugar specificity within the signature sequence and other residues that influence enzymatic activity are likely to be located distal to the signature sequence.

Plant UGTs are extremely diverse and have a broader range of UDP-sugar preferences than mammalian UGTs. Several studies involving plant UGTs have identified residues that contribute to sugar specificity, but none of these directly overlap with F391 in mammalian UGTs. Kubo et al (26) found that the Q within the DQ motif (Q394 in human UGT3A1) is conserved in plant and animal UDP glucuronosyltransferases and

glucosyltransferases, but that this residue is H in plant and animal UDP galactosyltransferases. Mutation of this residue from H to Q in a plant UDP galactosyltransferase lowered the Km of this enzyme for UDP-Glc approximately 40-fold suggesting that Q plays a role in glucose recognition. The complementary mutation of Q to H in a plant UDP glucosyltransferase impaired its ability to use UDP-Glc but did not confer activity with UDP-Gal (26). Moreover, mutation of the Q residue to H in another plant glucosyltransferase VvGT1 abolished activity (27). The D and Q residues were predicted to interact directly with the hydroxyl groups of the sugar moiety of UDP-Glc (27), a similar conclusion to that previously drawn for the DQ motif in binding of UDP-GlcUA (25).

Human UGT3A1 and UGT3A2 share the DQ (D393, Q394) motif with all other mammalian UGTs, suggesting that they are not important for sugar specificity. However, the two residues immediately after Q394 are divergent between UGT3A1 and other mammalian UGTs (see Figure 1). We found that mutation of these two residues in UGT3A1 impaired protein stability and activity with both UDP-GlcNAc and UDP-Glc (not shown). Thus the carboxyl-terminal end of the signature sequence may be involved indirectly in both sugar binding and catalysis by determining correct folding of UGT3A enzymes.

As well as identifying residues in UGT3A enzymes that confer specificity for UDP-GlcNAc or UDP-Glc, it is interesting to consider what sequence or structural features may preclude their use of UDP-GlcUA. Residues that confer UDP-GlcUA specificity have been examined using site-directed mutagenesis and protein modelling of BpUGT94B1, a UGT from red daisy (*Bellis perennis*) that conjugates glucuronic acid to flavonoids. This analysis identified an arginine outside of the signature sequence (R25) as crucial for activity with UDP-GlcUA, with R25 mutants exhibiting dramatically reduced activity with UDP-GlcUA, and 3-fold increased activity with UDP-Glc (28). A conserved arginine is found in the corresponding position relative to the mature N-terminus of all human UGTs of the UGT1 and UGT2 families. However, in UGT3A1 and UGT3A2 this residue is histidine (H49). This may support the notion that arginine at this position is important for utilization of UDP-GlcUA. It should be noted, however, that UGT8, which is only known to use UDP-Gal, also has an arginine at this position; thus the function of this residue is likely to be context dependent.

Is UDP-GlcNAc conjugation by UGT3A1 a primate innovation? Our cross-species genomic analysis indicates that N391-containing forms of the UGT3A family occur only in primates. Experimentally, we found that primate UGT3A1 proteins that contain the N391 residue (human, chimp and rhesus) were all capable of N-acetylglucosaminidation. In contrast, all of the UGT3A proteins from non-primates that we tested (cow, rabbit and mouse) lacked the N391 residue and were incapable of N-acetylglucosaminidation. Although it is not feasible to test all species, this sampling provides strong support for the idea of a primate specific UGT3A1 enzymatic activity. Additional circumstantial evidence for this notion is provided by the observations that 1) the signature sequences of primate UGT3A gene paralogues generally show lower sequence identity than non-primate paralogues, suggesting that the primate UGT3A enzymes are more functionally divergent, and 2) many non-primates have only one Ugt3a family gene (e.g. rat, rabbit, and elephant).

It is difficult to determine when the capacity for N-acetylglucosaminidation may have first emerged. Lemurs lack a N391-containing UGT3A isoform, suggesting that this residue may have first appeared around the time that haplorhini and strepsirrhini primates diverged 63 million years ago. However, both lemur genomes have lower-coverage (less than 2x) than the other primate genomes. Thus, future analyses of higher resolution genome assemblies may change this conclusion. There is also evidence for selective gene loss in some primates: e.g. gibbon and orang-utan have only one UGT3A gene (both of these genomes have 5-6x coverage). Similarly we have observed that the entire UGT3A family is absent in sequenced bird genomes although present in reptiles and some species of fish (unpublished observations).

Identification of endogenous substrates for human UGT3A enzymes as well as their developmental expression patterns may provide insight into the significance of primate diversification in sugar-conjugation capacity.

REFERENCES

1. Mackenzie, P. I., Owens, I. S., Burchell, B., Bock, K. W., Bairoch, A., Belanger, A., Fournel-Gigleux, S., Green, M., Hum, D. W., Iyanagi, T., Lancet, D., Louisot, P., Magdalou, J., Chowdhury, J. R., Ritter, J. K., Schachter, H., Tephly, T. R., Tipton, K. F., and Nebert, D. W. (1997) *Pharmacogenetics* **7**, 255-269
2. Meech, R., and Mackenzie, P. I. (1997) *Clin. Exp. Pharmacol. Physiol.* **24**, 907-915
3. Tukey, R. H., and Strassburg, C. P. (2000) *Annu. Rev. Pharmacol. Toxicol.* **40**, 581-616
4. Radominska-Pandya, A., Czernik, P. J., Little, J. M., Battaglia, E., and Mackenzie, P. I. (1999) *Drug Metab. Rev.* **31**, 817-899
5. Miners, J. O., and Mackenzie, P. I. (1991) *Pharmacol. Ther.* **51**, 347-369
6. Mackenzie, P. I., Bock, K. W., Burchell, B., Guillemette, C., Ikushiro, S.-i., Iyanagi, T., Miners, J. O., Owens, I. S., and Nebert, D. W. (2005) *Pharmacogenet. Genomics* **15**, 677-685
7. Ritter, J. K., Chen, F., Sheen, Y. Y., Tran, H. M., Kimura, S., Yeatman, M. T., and Owens, I. S. (1992) *J. Biol. Chem.* **267**, 3257-3261
8. Mackenzie, P. I., Rogers, A., Treloar, J., Jorgensen, B. R., Miners, J. O., and Meech, R. (2008) *J. Biol. Chem.* **283**, 36205-36210
9. Mackenzie, P. I., Rogers, A., Elliot, D. J., Chau, N., Hulin, J. A., Miners, J. O., and Meech, R. (2011) *Mol. Pharmacol.* **79**, 472-478
10. Mackenzie, P., Little, J. M., and Radominska-Pandya, A. (2003) *Biochem. Pharmacol.* **65**, 417-421
11. Tang, C., Hochman, J. H., Ma, B., Subramanian, R., and Vyas, K. P. (2003) *Drug Metab. Dispos.* **31**, 37-45
12. Senafi, S. B., Clarke, D. J., and Burchell, B. (1994) *Biochem. J.* **303**, 233-240
13. Bosio, A., Binczek, E., Le Beau, M. M., Fernald, A. A., and Stoffel, W. (1996) *Genomics* **34**, 69-75
14. Hobbs, S., Jitrapakdee, S., and Wallace, J. C. (1998) *Biochem. Biophys. Res. Commun.* **252**, 368-372
15. Udomuksorn, W., Elliot, D. J., Lewis, B. C., Mackenzie, P. I., Yoovathaworn, K., and Miners, J. O. (2007) *Pharmacogenet. Genomics* **17**, 1017-1029
16. Eddy, S. R. (1998) *Bioinformatics* **14**, 755-763
17. Radominska-Pandya, A., Bratton, S. M., Redinbo, M. R., and Miley, M. J. (2010) *Drug Metab. Rev.* **42**, 128-139
18. Powell, M. J. D. (1977) *Mathematical Programming* **12**, 241-254
19. Mayo, S. L., Olafson, B. D., and Goddard, W. A. (1990) *The Journal of Physical Chemistry* **94**, 8897-8909
20. Bowalgaha, K., Elliot, D. J., Mackenzie, P. I., Knights, K. M., Swedmark, S., and Miners, J. O. (2005) *Br. J. Clin. Pharmacol.* **60**, 423-433
21. Meech, R., and Mackenzie, P. I. (2010) *Drug Metab. Rev.* **42**, 43-52
22. Miley, M. J., Zielinska, A. K., Keenan, J. E., Bratton, S. M., Radominska-Pandya, A., and Redinbo, M. R. (2007) *J. Mol. Biol.* **369**, 498-511
23. Battaglia, E., Terrier, N., Mizeracka, M., Senay, C., Magdalou, J., Fournel-Gigleux, S., and Radominska-Pandya, A. (1998) *Drug Metab. Dispos.* **26**, 812-817
24. Patana, A.-S., Kurkela, M., Goldman, A., and Finel, M. (2007) *Mol. Pharmacol.* **72**, 604-611

25. Xiong, Y., Patana, A.-S., Miley, M. J., Zielinska, A. K., Bratton, S. M., Miller, G. P., Goldman, A., Finel, M., Redinbo, M. R., and Radominska-Pandya, A. (2008) *Drug Metab. Dispos.* **36**, 517-522
26. Kubo, A., Arai, Y., Nagashima, S., and Yoshikawa, T. (2004) *Arch. Biochem. Biophys.* **429**, 198-203
27. Offen, W., Martinez-Fleites, C., Yang, M., Kiat-Lim, E., Davis, B. G., Tarling, C. A., Ford, C. M., Bowles, D. J., and Davies, G. J. (2006) *EMBO J.* **25**, 1396-1405
28. Osmani, S. A., Bak, S., Imberty, A., Olsen, C. E., and Moller, B. L. (2008) *Plant Physiol.* **148**, 1295-1308. Epub 2008 Oct 1291.

FOOTNOTES

This work was supported by the National Health and Medical Research Council of Australia. PIM is a NHMRC Senior Principal Research Fellow. We thank Ms. Julie-Ann Hulin for technical assistance.

The abbreviations used are: HEK, human embryonic kidney; HDCA, hyodeoxycholic acid; MCA, mycophenolic acid; 4-MU, 4-methylumbelliferone; RT-PCR, reverse transcriptase polymerase chain reaction; UDCA, ursodeoxycholic acid; UGT, UDP glycosyltransferase; UDP-GlcNAc, UDP N-acetylglucosamine; UDP-GlcUA, UDP glucuronic acid; UDP-Glc, UDP glucose; UDP-Gal; UDP galactose; UDP-Xyl, UDP xylose.

FIGURE LEGENDS

Figure 1. Alignment of the UGT signature sequences of all human UGTs: the residue at position 391 is divergent in human UGT3A1.

Figure 2. Docked UDP-GlcNAc (sticks) positioned for productive S_N2 nucleophilic attack. Stabilization of the uridine, diphosphate, and sugar substituents of the cofactor involves hydrogen bonding interactions with highly conserved residues of the UGT signature sequence (cartoon). Bonding interactions are observed from L352/Q354/E377 to the uridine, Q372/N373/S374 to the diphosphate, and H369/Q394 to the sugar moiety.

Figure 3. Energetically favoured poses of cofactor after automated *in silico* docking. (A) UDP-GlcNAc with wild-type UGT3A1, (B) UDP-Glc with wild-type UGT3A1, (C) UDP-GlcNAc with UGT3A1(N391F) mutant, and (D) UDP-Glc with UGT3A1(N391F) mutant

Figure 4. Alignment of UGT3A signature sequences from multiple primate and non-primate species. Genomes were accessed using the NCBI and Ensembl genome browsers. Accession numbers are given for each database where the gene had been annotated. BLAST analysis was also used to identify all possible homologues within each locus. Pairwise comparisons were performed using CLUSTALW and the percentage identity between the signature sequences of paralogous genes is given to the right of the table. For dog the percentage identity is given as the mean of the three pairwise comparisons.

Figure 5. A) Analysis of *Ugt3a1* and *Ugt3a2* mRNA levels in neonatal mouse liver and kidney by quantitative RT-PCR. Data were analysed using a standard curve for each gene and normalized to the housekeeping gene RPS26. B) Analysis of mouse Ugt3a2 enzymatic activity with its preferred aglycone substrate hyodeoxycholic acid (HDCA) and either UDP-Glc or UDP-GlcNAc as sugar donor. Assays were chromatographed and visualized by autoradiography. The arrow indicates the HDCA glucoside,

the asterisk indicates endogenous glycosidation activities in HEK293T cells. n.c. – negative control HEK293T lysate. C) Analysis of endogenous glucosidation and N-acetylglucosaminidation activities in mouse liver and kidney microsomal preparations. Assays were performed using a range of aglycone substrates and either UDP-Glc or UDP-GlcNAc as sugar donor, and analysed as in B).

Figure 6. Analysis of the enzymatic activities of various primate and non-primate UGT3A isoforms using preferred aglycone substrates and either UDP-Glc or UDP-GlcNAc as sugar donor. Assays were chromatographed and the glycosides products visualized by autoradiography. Human UGT3A1 and 3A2 served as positive controls.

Table 1. Activities of wildtype and mutated forms of human UGT3A1 and UGT3A2 with genistein and either UDP-Glc or UDP-GlcNAc as sugar donor.

Protein variant	**Activity with UDP-Glc (pmol/min/mg)**	**Activity with UDP-GlcNAc (pmol/min/mg)**
UGT3A1 wt	1.75 +/- 0.05	65.74 +/- 13.13
UGT3A1 N391F	15.03 +/- 1.33	nd
UGT3A2 wt	137.65 +/- 6.59	nd
UGT3A2 F391N	0.82 +/- 0.08	nd

nd = not detected

Table 2. Activities of wildtype mouse Ugt3a1 and Ugt3a2 with various aglycone substrates and either UDP-Glc or UDP-GlcNAc as sugar donor.

Protein variant	**Activity with UDP-Glc (pmol/min/mg)**		**Activity with UDP-GlcNAc (pmol/min/mg)**	
	4MU	**UDCA**	**4MU**	**UDCA**
mUGT3A1	3.08	5.56	nd	nd
mUGT3A2	4.77	2.84	nd	nd

nd = not detected

Figure 1


hUGT3A1 WLPQSDLLAHPSIRLFVTHGGQNSVMETIRHGVPMVGLPVNGDQHG
hUGT3A2 WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQPE
hUGT1A1 WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMD
hUGT1A3 WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMD
hUGT1A4 WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMD
hUGT1A5 WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMD
hUGT1A6 WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMD
hUGT1A7 WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMD
hUGT1A8 WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMD
hUGT1A9 WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMD
hUGT1A10 WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMD
hUGT2A1 WIPQNDLLGHPKTKAFITHGGTNGIYEAIYHGVPMVGVPMFADQPD
hUGT2A3 WIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLD
hUGT2B7 WIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFADQPD
hUGT2B28 WIPQNDLLGLPKTRAFITHGGANGIYEAIYHGIPMVGIPLFWDQPD
hUGT2B10 WIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFFDQPD
hUGT2B11 WIPQNDLLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFFDQPD
hUGT2B15 WLPQNDLLGHPKTKAFITHGGTNGIYEAIYHGIPMVGIPLFADQDD
hUGT2B17 WLPQNDLLGHPKTKAFITHGGTNGIYEAIYHGIPMVGIPLFADQHD
hUGT8 WLPQNDLLGHSKIKAFLSHGGLNSIFETMYHGVPVVGIPLFGDHYD

Figure 2

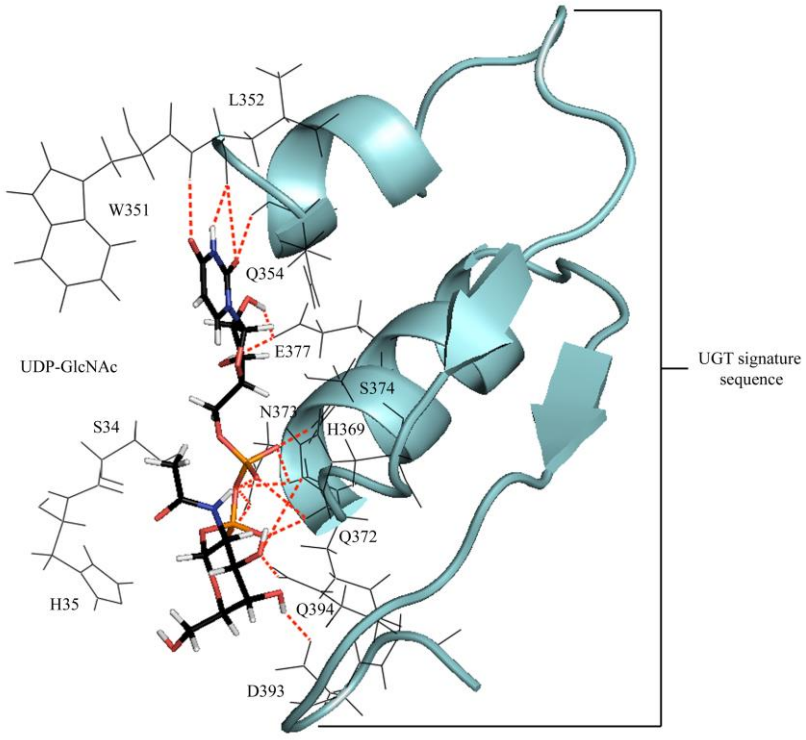

L352
W351
Q354
E377
UDP-GlcNAc
S374
N373
H369
S34
Q372
H35
Q394
D393
UGT signature
sequence

Figure 3

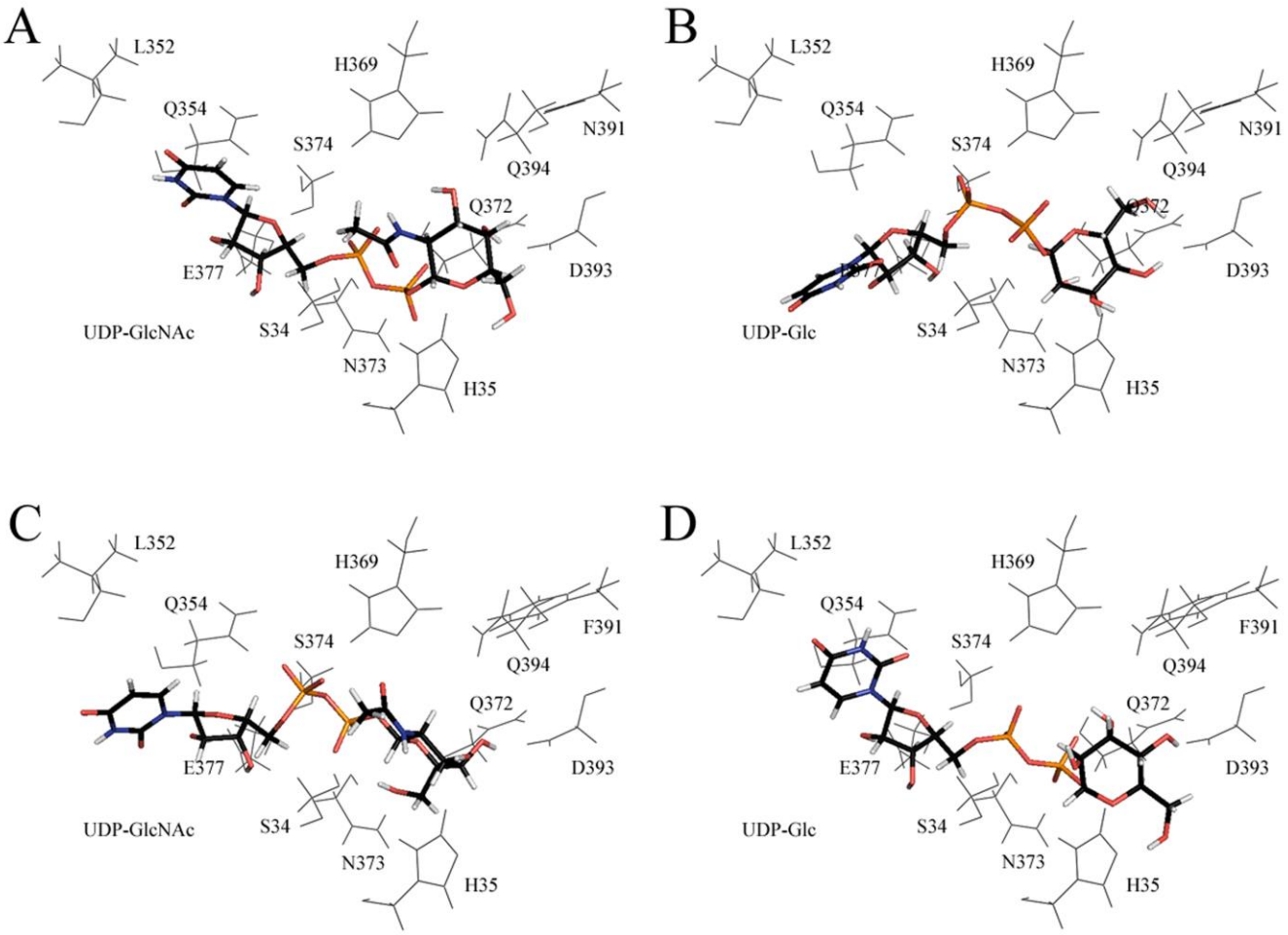

A
L352
H369
Q354
S374
N391
Q394
Q372
E377
D393
UDP-GlcNAc
S34
N373
H35
B
L352
H369
Q354
S374
N391
Q394
Q372
D393
UDP-Glc
S34
N373
H35
C
L352
H369
Q354
S374
F391
Q394
Q372
E377
D393
UDP-GlcNAc
S34
N373
H35
D
L352
H369
Q354
S374
F391
Q394
Q372
E377
D393
UDP-Glc
S34
N373
H35

Figure 4

Species/gene orthologue	NCBI Gene ID#	Ensembl Accession	Signature sequence	
Primates				
human 3A1	133688	ENSG00000145626	WLPQSDLLAHPSIRLFVTHGGQNSVMETIRHGVPMVGLPV**N**GDQHG	82%
human 3A2	167127	ENSG00000168671	WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQPE	
chimp 3A1	471571	ENSPTRG00000016792	WLPQSDLLAHPSIRLFVTHGGQNSVMEAIRHGVPMVGLPV**N**GDQHG	84%
chimp 3A2	461920		WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQPE	
gorilla 3A1		ENSGGOG00000008662	WLPQSDLLAHPSIRLFVTHGGQNSVMEAIRHGVPMVGLPV**N**GDQHG	82%
gorilla 3A2		ENSGGOG00000003108	WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPVVGIPLFGDQPE	
orang-utan 3A1	100174402	ENSPPYG00000015398	WLPQSDLLAHPSICLFVTHGGQNSVMEAIRHGVPMVGLPV**N**GDQHG	
baboon 3A1		ENSP00000274278	WLPQSDLLAHPSIRLFVTHGGQNSVMEAIWHGVPMVGLPV**N**GDQHG	84%
baboon 3A2		ENSP00000282507	WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQPE	
macaque 3A1	700115	ENSMMUG00000006907	WLPQSDLLAHPSIRLFVTHGGQNSVMEAIRHGVPMVGLPV**N**GDQHG	84%
macaque 3A2	700238	ENSMMUG00000006910	WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQPE	
marmoset 3A1	100399428	ENSCJAG00000019787	WLPQSDLLAHPSIRLFVTHGGQNSIMEAIRHGVPMVGLPV**N**GDQHG	89%
marmoset 3A2	100396365	ENSCJAG00000018354	WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPVFGDQPE	
tarsier 3A1		ENSTSYG00000006598	WFPQSDLLAHPNIHLFVTHGGQNSIMETIQYGVPIVGIPV**N**GDQHD	76%
tarsier 3A2		ENSTSYG00000004761	WLPQRDLLAHPSIRLLVTHGGQNSIMEAIQHGVPMVGIPVFGDQPE	
gibbon 3A2	100584976	ENSNLEG00000015982	WLPQSDLLAHPSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQPE	
Bushbaby (otolemur) 3A2		ENSOGAG00000007228	WLPQSDLLAHPRIRLFVTHGGQNSIMEAIQHGVPLVGIPVFGDQSE	
mouse lemur 3A2		ENSMICG00000006575	WLPQSDLLAHPSIRLFVTHGGQNSIMEAIRHGVPMVGIPVFGDQLE	
Non-primates*				
mouse "3A1"	105887	ENSMUSG00000072664	WLPQIDLLAHPSIRLFVTHGGMNSVMEAVHHGVPMVGIPFFGDQPE	95%
mouse "3A2"	223337	ENSMUSG00000049152	WLPQTDLLAHPSIRLFVTHGGMNSVMEAVHHGVPMVGIPFFFDQPE	
rat 3A**	294793	ENSRNOG00000025010	WLPQTDLLAHPSIRLFVTHGGMNSVNEAIQHGVPMVGILFFSDQPE	
rabbit 3A**	100353356	ENSOCUG00000011414	WLPQNDLLAHPNIRLFVTHGGQNSIMEAIQYGVPMVGIPLFGDQPE	
dog"3A2"	489230	ENSCAFG00000023144	WLPQNDLLAHPHIRLFVTHGGMNSIMEAIQHGVPMVGIPLFGDQPE	
dog "3A1-like"	489231	ENSCAFG00000018738	WLPQNDLLAHPRIRLFVTHGGMNSIMEAIQHGVPMVGIPVLGDQPE	91%
dog "3A1"	489234	ENSCAFG00000018743	WLPQNDLLAHPHIRLFVTHGGMNSIMEAIQHGVPMIGIPVFGEQAE	
cow "3A1"	525484		WLPQNDLLAHSHIRLFVTHGGMNSIMEAIHHGVPMVGIPVFEDQDE	82%
cow "3A2"	537188	ENSBTAG00000002701	WLPQNDLLGHPRIRLFVSHGGMNSIMEAIQHGVPMVGIPLFGDQHE	
horse "3A1-like"	100067792	ENSECAG00000010396	WLPQSDLLAHPRIRLFVTHGGINSIMEAIQHGVPMVGIPFFGDQPE	95%
horse "3A1-like"	100067844		WLPQSDLLAHPRIRLFVTHGGLNSIMEAIQHGVPMVGIPLFGDQPE	
panda "3A1-like"	100477988	ENSAMEG00000012397	WLPQNDLLAHPRIRLFVTHGGMNSIMEAIQHGIPVVGIPVLGDQPD	84%
panda "3A2-like"	100477732		WLPQSDLLAHPCIRLFVTHGGMNSIMEAIQHGVPMVGIPLFGDQPE	
Elephant 3A**	100665586		WLPQSDLLAHSSIRLFVTHGGQNSIMEAIQHGVPMVGIPLFGDQPE	
platypus "3A2-like"	100087188		WLPQNDLLAHPKAKLLVTHGGINSVMEAIHHGVPMVGIPLFGDQFD	

* In these species the homologues of human UGT3A1 and UGT3A2 cannot generally be determined by sequence analysis alone, thus the names are placed in inverted quotes. ** Rat, rabbit, and elephant have only one UGT3A form.

Figure 5

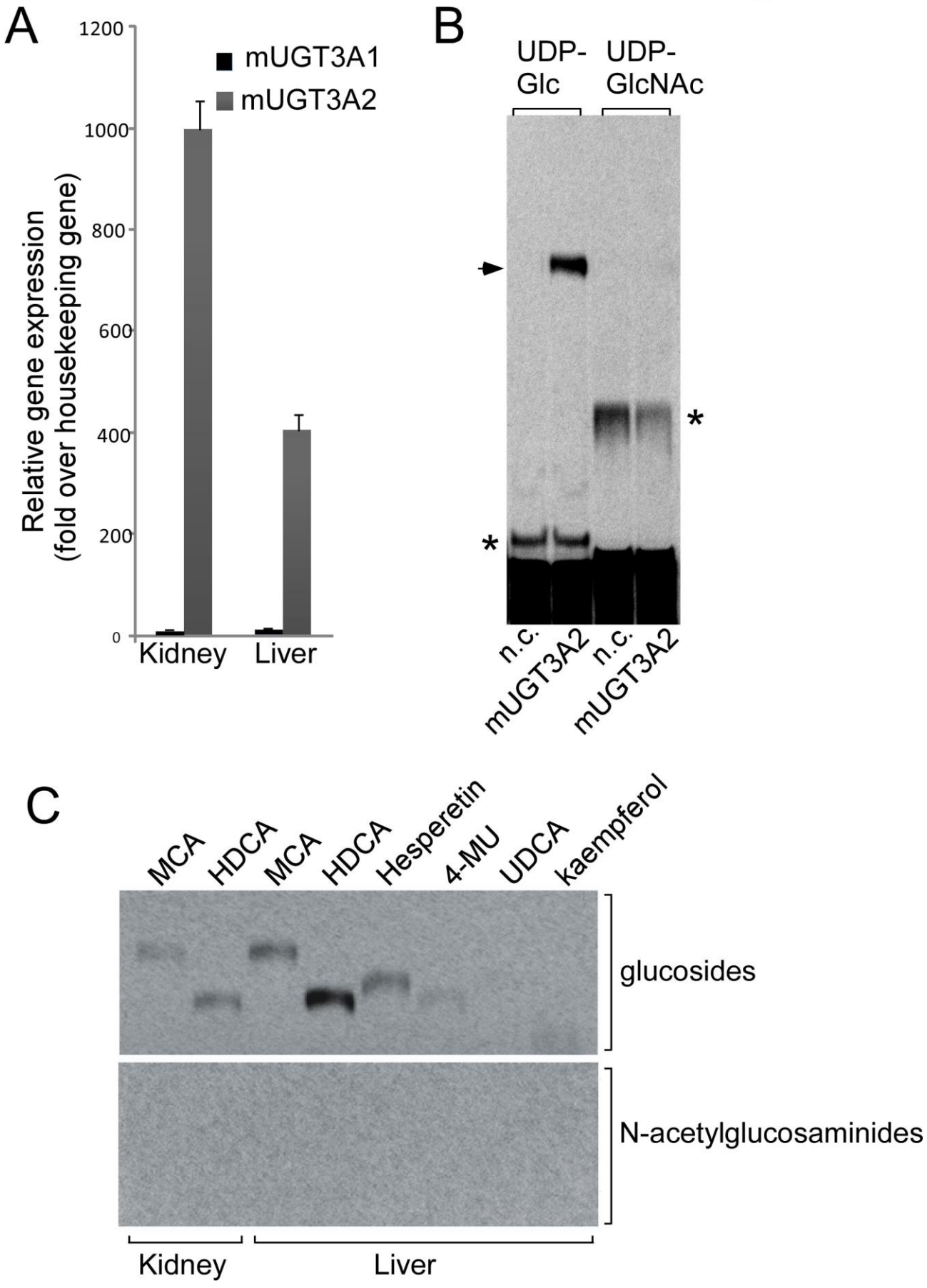
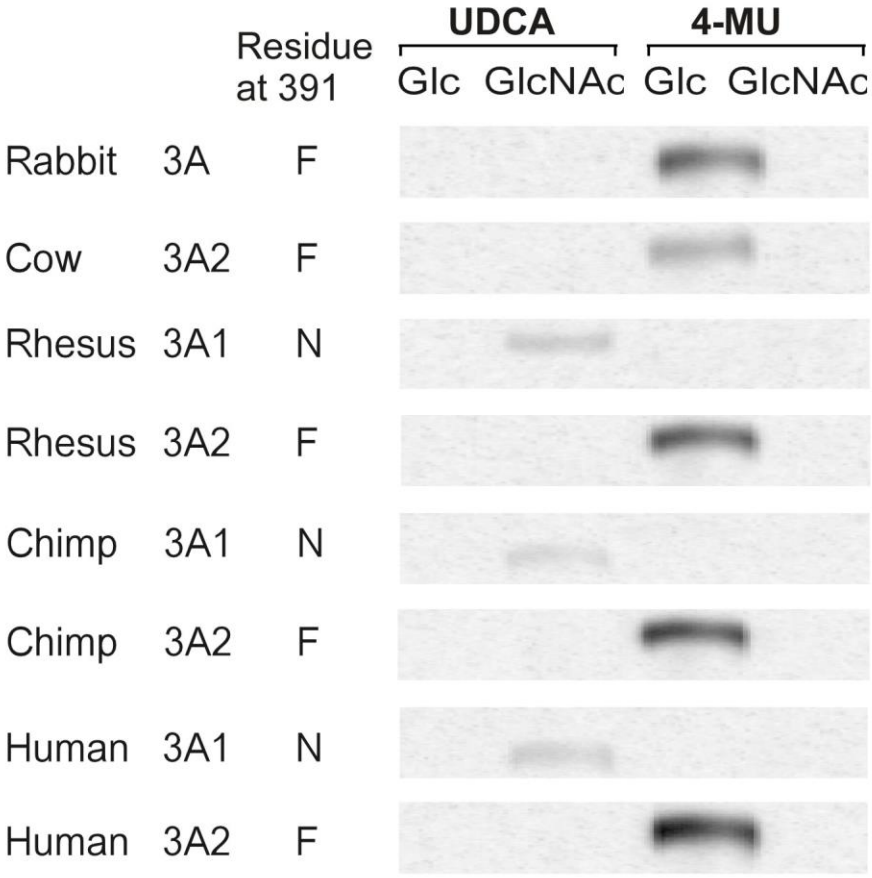

A
1200
1000
800
600
400
200
0
mUGT3A1
mUGT3A2
Relative gene expression
(fold over housekeeping gene)
Kidney
Liver
B
UDP-
Glc
UDP-
GlcNAc
*
*
n.c.
mUGT3A2
n.c.
mUGT3A2
C
MCA
HDCA
MCA
HDCA
Hesperetin
4-MU
UDCA
kaempferol
glucosides
N-acetylglucosaminides
Kidney
Liver

Figure 6


Residue at 391
UDCA
4-MU
Glc
GlcNAc
Glc
GlcNAc
Rabbit 3A F
Cow 3A2 F
Rhesus 3A1 N
Rhesus 3A2 F
Chimp 3A1 N
Chimp 3A2 F
Human 3A1 N
Human 3A2 F