The Importance of Local Chemical Structure for Chemical Metabolism by Human Uridine 5′-Diphosphate−Glucuronosyltransferase

Michael J. Sorich,*,† Ross A. McKinnon,† John O. Miners,‡ and Paul A. Smith‡

Sansom Institute, School of Pharmacy and Medical Sciences, University of South Australia, Frome Rd, Adelaide, SA 5000, Australia, and Department of Clinical Pharmacology, Flinders University, Bedford Park, SA 5042, Australia

Received June 20, 2006

The uridine 5′-diphosphate− (UDP−)glucuronosyltransferase (UGT) family of enzymes catalyzes the conjugation of chemicals containing a suitable nucleophilic atom with glucuronic acid. Despite the importance of glucuronidation as an elimination and detoxification mechanism for drugs, environmental chemicals, and endogenous compounds, the structural features of substrates that confer isoform selectivity are poorly understood. The relationship between the local molecular structure of nucleophilic atoms of chemicals and the ability of UGT isoforms to glucuronidate the nucleophilic atoms was investigated here. The proximity of an aromatic ring to the nucleophilic atom was highly associated with a greater likelihood of glucuronidation by most UGT isoforms. Similarly, most UGT isoforms were found to have a statistically significant preference for oxygen over nitrogen as the nucleophilic atom. The converse was established only for UGT1A4. Naïve Bayes models were trained to predict the site of glucuronidation for eight UGT isoforms on the basis of the partial charge and Fukui function of the nucleophilic atom and whether an aromatic ring was attached to the nucleophilic atom. On average, the cross-validated sensitivity and specificity of the models were approximately 75−80%. For all but UGT2B7, the area under the receiver operating characteristics curve of the model was greater than 0.8, indicating strong predictive ability. A chemical diversity analysis of the currently available data indicates bias toward chemicals containing phenolic groups, and it is likely that the availability of chemical data sets with greater diversity will facilitate further insights into the structural features of substrates that confer enzyme selectivity.

INTRODUCTION

Humans are exposed on a daily basis to a diverse array of foreign chemicals (xenobiotics), including pharmaceutical, environmental, and industrial chemicals. The detoxification of xenobiotics within the human body relies heavily on chemical modification (metabolism) by a relatively small number of enzymes.[1] Metabolism promotes excretion of the chemical via the kidneys or bile and, generally, reduces pharmacological and toxicological activity. While most enzymes only interact with a small set of structurally similar chemicals, some chemical detoxification enzymes have evolved the ability to metabolize chemicals of highly divergent sizes, shapes, flexibilities, complexities, and hydrophobicities (e.g., Figure 1).[1] However, the structural features of substrates that confer enzyme selectivity are frequently poorly understood.

The uridine 5′-diphosphate− (UDP−)glucuronosyltransferase (UGT) family of enzymes comprises 15 functional "isoforms" that catalyze the conjugation of chemicals containing a suitable nucleophilic atom (mainly, aliphatic or aromatic hydroxyl, carboxyl, or amino groups) with glucuronic acid, according to a second-order nucleophilic substitution mechanism.[2−4] Glucuronidation serves as a clearance mechanism for drugs from all therapeutic classes, including some with a narrow therapeutic index.[2−10] In addition, glucuronidation acts as an elimination pathway for numerous other endogenous, dietary, and environmental chemicals.[4]

A single UGT isoform can catalyze the glucuronidation of multiple "acceptor" sites that may be present within a single compound,[5,6,8,10] suggestive of multiple substrate binding modes. Furthermore, the kinetic modeling of xenobiotic glucuronidation is consistent with multiple substrate binding sites within the UGT active site.[11,12] UGT is a membrane-bound protein with a poorly defined three-dimensional structure (the most similar protein with a solved high-resolution crystal structure shares approximately 15% sequence identity). Consequently, insights into the molecular mechanism underlying substrate recognition are predominantly based on small molecule modeling.[3,13−23]

Previously, in silico models were made to predict substrates and nonsubstrates of human UGT isoforms ("reaction phenotyping").[22] It was demonstrated that both 2D molecular descriptors and quantum-chemical molecular descriptors were useful for discriminating substrates from nonsubstrates.[20,22] Multiple pharmacophore perception studies gave insight into the chemical features associated with substrates and nonsubstrates of human UGT isoforms.[21] The individual pharmacophores differed between isoforms but, generally, represented relatively simple structural and chemical features.[21] Furthermore, the pharmacophores suggested that the environ-

* Corresponding author tel.: +61 8 83021230; fax: +61 8 83021087; e-mail: michael.sorich@unisa.edu.au.
† University of South Australia.
‡ Flinders University.

10.1021/ci600248e CCC: $33.50
Published on Web 09/22/2006

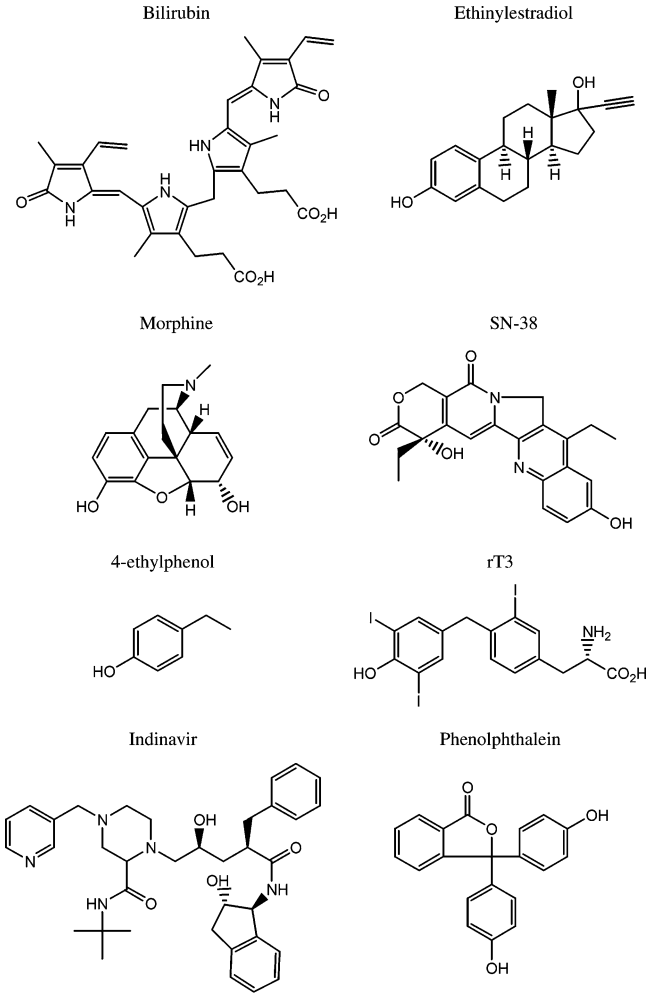

Figure 1. Examples of structurally diverse chemicals that bind to UGT1A1.

ment immediately adjacent to the nucleophilic site of conjugation may be an important determinant of metabolism by a particular UGT.[21] However, because of a lack of data on actual sites on glucuronidation on the substrates, this was not able to be confirmed.

To gain a greater insight into the mechanisms underlying glucuronidation, this study seeks to understand the features of nucleophilic atoms of chemicals that effect whether a reaction occurs at specific nucleophilic atoms or not.

MATERIALS AND METHODS

Data Sets. Potential sites of glucuronidation on a chemical include the following nucleophilic atoms: oxygen in a hydroxyl group, singly bonded oxygen in a carboxylic acid group, nitrogen in an amine group or in a heterocyclic ring, and sulfur in a thiol group. For each UGT isoform, a data set was collated from the literature on nucleophilic atoms which are glucuronidated and nucleophilic atoms which are not glucuronidated (see the Supporting Information). A nucleophilic atom is deemed to be glucuronidated if (1) the specific position of the glucuronide was experimentally determined following an assay of the chemical in an in vitro recombinant cell system expressing a single UGT isoform or (2) the chemical contains only one nucleophilic atom and the chemical was determined to be glucuronidated in an in vitro recombinant cell system expressing a single UGT isoform. Similarly, a nucleophilic atom was deemed not to be glucuronidated if (1) the glucuronide could not be detected following an assay of the chemical in an in vitro recombinant cell system expressing a single UGT isoform or (2) the chemical was found not to be glucuronidated at all in an in vitro recombinant cell system expressing a single UGT isoform.

Hierarchical Clustering. For each UGT isoform, the nucleophilic atoms in the data set were assigned to one or more functional group types: phenolic (Ar−OH), aliphatic hydroxyl (Al−OH), carboxylic acid (COOH), amine adjacent to aromatic (Ar−NR_x) aliphatic amine (Al−NR_x), primary amine (NR_1), secondary amine (NR_2), tertiary amine (NR_3), heterocyclic aromatic nitrogen (n), and thiol (SH). For each functional group type with more than 10 known nucleophilic atoms assigned for a particular UGT isoform, the percentage of nucleophilic atoms glucuronidated was calculated. If insufficient data existed, the percentage was labeled "NA".

To display the similarity of nucleophilic atom preference between UGT isoforms and between functional group types, the percentage of each functional group type glucuronidated by individual UGT isoforms was hierarchically clustered using the Euclidean distance with "NA" data masked.[24] The percentage of each local structure glucuronidated by each UGT isoform is visually represented as a color map, using linear interpolation between the dark blue and light blue representing 100% glucuronidation and 0% glucuronidation, respectively.

Statistical Tests of Local Chemical Properties. When the R statistical environment was used,[25] the Fisher exact test[26] was used to determine the statistical significance of the effect of two local features on the likelihood of nucleophilic atom glucuronidation. For each UGT isoform, the count of glucuronidated and nonglucuronidated nucleophilic atoms was determined for (1) nucleophilic atoms attached to an aromatic ring versus those not attached to an aromatic ring and (2) nucleophilic nitrogen atoms (e.g., all amines and aromatic nitrogens) versus nucleophilic oxygen atoms (e.g., all hydroxyls and carboxylic acids).

The partial charge and Fukui function were calculated using an in-house implementation of the electronegativity equalization method.[27] The nonparametric Mann−Whitney[28] test implemented in R[25] was used to determine whether the partial charge and Fukui function of the nucleophilic atom were useful for discriminating glucuronidation reactivity.

Naïve Bayes Models. To determine the importance of the local structure for the prediction of glucuronidation of the nucleophilic atom, a naïve Bayes model[29] (implemented in the Orange data mining software[30]) was trained for each UGT isoform for which at least 20 examples existed of nucleophilic atoms that are and are not glucuronidated. Naïve Bayes classification was chosen for its simplicity and ability to handle unbalanced data sets.[29] Three explanatory variables were used: a Boolean variable indicating whether an aromatic ring was attached to the nucleophilic atom, the partial charge of the nucleophilic atom, and the atom-centered Fukui function. To produce models with similar classification performance for glucuronidated and nonglucuronidated nucleophilic atoms, weighting of the two groups was set to counteract the imbalance in the training set sizes of the two groups. The predictive ability of the models was assessed using 10-fold cross-validation with stratification of the

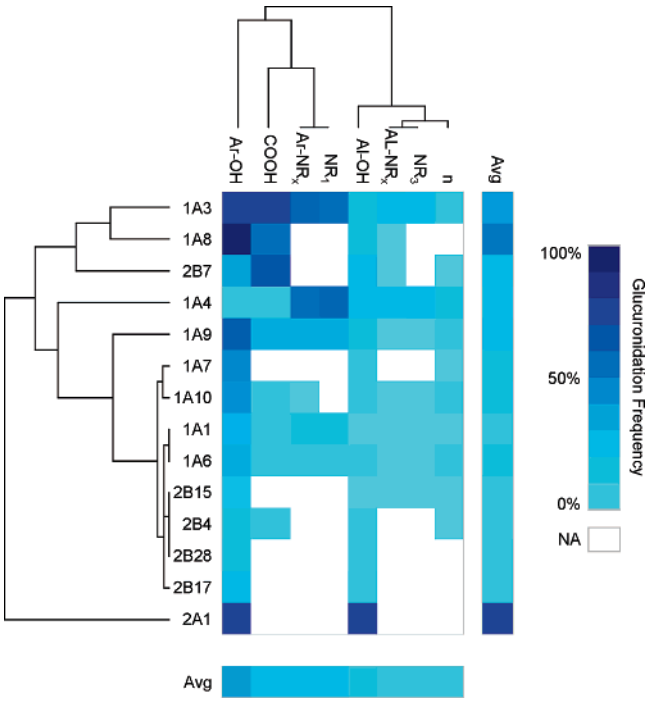

Figure 2. Clustering and color representation of glucuronidation frequency of different nucleophile types by 14 UGT isoforms (1A1, 1A3, 1A4, 1A6, 1A7, 1A8, 1A9, 1A10, 2A1, 2B4, 2B7, 2B15, 2B17, and 2B28). The nucleophile types displayed include the following functional groups: phenolic (Ar−OH), aliphatic hydroxyl (Al−OH), carboxylic acid (COOH), amine adjacent to aromatic (Ar−NR_x), aliphatic amine (Al−NR_x), primary amine (NR_1), tertiary amine (NR_3), and aromatic nitrogen (n).

samples on the basis of class.[31] This assessment method is generally equivalent to using a random test set consisting of 10% of the data, but with the process of selecting and assessing the test set repeated 10 times and averaged (see ref 31 for a review on cross-validation). As a result, the test set performance is not so sensitive to the selection of the test set. Sensitivity, specificity, and the area under the receiver operating characteristic curve (AROC)[32] were subsequently used to summarize the predictive ability of the model.

Comparison of Chemical Libraries. The set of approximately 500 chemicals that have been tested for glucuronidation by individual UGT isoforms was compared against the NCI diversity set[33] of approximately 2000 chemicals and the 3D Pharmaceutical Structure Database (http://www.ps.toyaku.ac.jp/dobashi/) of approximately 1000 drugs. The number of chemicals that contained each of the functional group types was ascertained using SMARTS substructure searching implemented in the Frowns toolkit (http://frowns.sourceforge.net/). The frequency of a functional group was defined as the number of chemicals that contain the functional group divided by the total number of chemicals in the specific chemical library.

RESULTS AND DISCUSSION

The color map in Figure 2 displays the glucuronidation frequency of nucleophilic atoms with differing local structures and substitutions. The classifications are based on standard functional-group types, as this is an intuitive and informative way to represent both the local electronic and steric environment of the nucleophilic atom. It is immediately apparent from the color map that differences exist in the susceptibility of different nucleophile types to glucuronidation, indicating that the local environment of the nucleophilic atom has an important influence on the potential for glucuronidation.

One striking feature of Figure 2 is the difference in susceptibility to glucuronidation between aliphatic (Al−OH and Al−NR_x) and aromatic (Ar−OH and Ar−NR_x) nucleophilic groups. For almost all UGT isoforms, it appears that an aromatic ring attached to the nucleophilic atom increases, often dramatically, the susceptibility of the nucleophilic atom to glucuronidation. This observation was subsequently statistically assessed using the Fisher exact test and Mann−Whitney test (Table 1). For most of the isoforms for which sufficient data exist, the differences between nucleophilic atoms attached to an aromatic ring and those not attached to an aromatic ring were highly statistically significant. Differences were nonsignificant only for UGT1A4 and UGT2B7. It is possible that the aromatic ring increases the likelihood of glucuronidation at the nucleophilic atom by enhancing the percentage of productive substrate-binding events within the active site, perhaps by forming π-stacking interactions with aromatic ring(s) in the active site.[34] Such an interaction may align the nucleophilic atom more appropriately for reaction with the UDP−glucuronic acid cofactor and augment the reaction rate by increasing the frequency of successful bond forming or by decreasing the entropy of bond formation.[34] It is also possible that the aromatic ring may increase the reaction rate by electronically stabilizing the reaction intermediate.

Figure 2 also indicates that, within analogous environments, nucleophilic oxygen atoms may be preferred for glucuronidation over nucleophilic nitrogen atoms. This observation was also verified statistically by determining the probability that differences in susceptibility to glucuronidation between nucleophilic nitrogen and oxygen atoms can be accounted for by chance (Table 1). For many isoforms, the probability of these differences occurring by chance is extremely unlikely ($p < 0.001$). UGT1A4 is the exception to the rule, showing a preference for glucuronidation of nucleophilic nitrogens over oxygens. For all UGT isoforms analyzed, the partial charge and the Fukui function were on average more negative and higher in nucleophilic atoms that are glucuronidated, respectively. In almost all cases, the differences were highly statistically significant. The association of negative charge and the Fukui function with glucuronidation is probably because they indicate greater nucleophilicity. The Fukui function indicates how incoming or outgoing electrons are redistributed in various regions of the molecule. As such, the atom-condensed Fukui function (the Fukui function integrated over the region attributed to an atom) is useful as an indicator of the relative atomic susceptibility to electrophilic or nucleophilic attack.[35−37]

Naïve Bayes classifiers were developed in order to estimate how well the glucuronidation of nucleophilic atoms could be predicted on the basis of the local environment of the nucleophilic atom. Such models predict the glucuronidation of nucleophilic atoms by each UGT isoform using only three simple descriptors of the nucleophilic atom's local structure. The percentage of glucuronidated and nonglucuronidated nucleophilic atoms predicted correctly, as determined by 10-fold cross-validation, is generally high (Table 1). The median

Table 1. Data Set Size, Association with Descriptors, and Multivariate Model Results for each UGT isoform

	number of nucleophilic sites		*p* value (Fisher exact or Mann−Whitney test)				naïve Bayes model (based on charge, Fukui, and arom[a]) cross-validation results		
isoform	glucuronidated	not glucuronidated	OvsN[b]	arom[a]	charge[c]	Fukui[d]	% sensitivity	% specificity	AROC[e]
UGT1A1	36	253	0.009	$<10^{-7}$	>0.05	$<10^{-3}$	81	77	0.81
UGT1A3	72	104	0.002	$<10^{-8}$	$<10^{-4}$	$<10^{-3}$	87	76	0.86
UGT1A4	54	201	0.017	>0.05	$<10^{-5}$	0.002	86	78	0.86
UGT1A6	54	267	$<10^{-3}$	$<10^{-8}$	$<10^{-4}$	$<10^{-6}$	78	70	0.86
UGT1A8	51	45	0.002	$<10^{-7}$	>0.05	$<10^{-5}$	80	88	0.92
UGT1A9	81	184	$<10^{-6}$	$<10^{-12}$	$<10^{-8}$	$<10^{-8}$	79	86	0.90
UGT1A10	47	183	$<10^{-4}$	$<10^{-9}$	0.009	$<10^{-4}$	84	73	0.86
UGT2B7	59	150	$<10^{-5}$	>0.05	$<10^{-3}$	0.008	79	56	0.70

[a] arom = Boolean variable indicating whether an aromatic ring is attached to a nucleophilic atom. [b] OvsN = Boolean variable indicating whether the nucleophilic atom is oxygen or nitrogen. [c] charge = Atom-centered partial charge. [d] Fukui = Atom-centered Fukui function. [e] AROC = Area under the receiver operating characteristic curve.

sensitivity, specificity, and AROC are 81%, 76%, and 0.86, respectively. These results demonstrate that the local structure of the nucleophilic atom is a very important predictor of glucuronidation and, in many cases, is the major predictor of glucuronidation (random guess = 50% correct; best model possible given the experimental error in data ~ 90% correct). This is striking considering the wide diversity in chemical size, shape, flexibility, and hydrophobicity and provides some insight into how UGT isoforms are able to glucuronidate such an extensive range of chemicals. If UGTs predominantly recognize such local structures, this could facilitate the catalysis of chemicals with greatly divergent global structures. Molecular recognition is complex (especially for enzymes such as UGT with multiple binding modes or sites), and while there are obviously many additional factors that influence whether specific nucleophilic atoms will be glucuronidated by UGTs, the evidence presented here suggests that the local environment of nucleophilic atoms is of great importance. It is likely that higher-quality quantitative data regarding glucuronidation at nucleophilic atoms will be useful in gaining a deeper insight into how nucleophilic atom local structure influences the glucuronidation capacity. It should be noted, however, that there are significant experimental issues (e.g., standardization of methods, quantification of isoform protein concentration, and high-throughput chemical quantification methods) that must be overcome, before the generation of such a data set becomes likely.[4]

The white patches in Figure 2 indicate that very limited data are available for the glucuronidation of certain nucleophile types by certain UGT isoforms. This is especially apparent for the UGT2B family. Furthermore, the susceptibility to glucuronidation of thiols and secondary amines could not be estimated for any UGT isoform, and frequently, the likelihood of amine (and to a lesser extent carboxylic acid) glucuronidation could not be estimated. By far, the most data exist for glucuronidation at aliphatic and aromatic hydroxyls. To assess the significance of this imbalance in the context of "chemical diversity space", the set of all chemicals tested for glucuronidation in an assay containing any individual UGT isoform was compared against (i) approximately 1000 marketed drugs and (ii) the NCI diversity set containing approximately 2000 diverse chemicals (Figure 3).

In general, the NCI data set had the most uniform distribution of functional group types. This is to be expected, considering that the NCI chemical set has been selected for

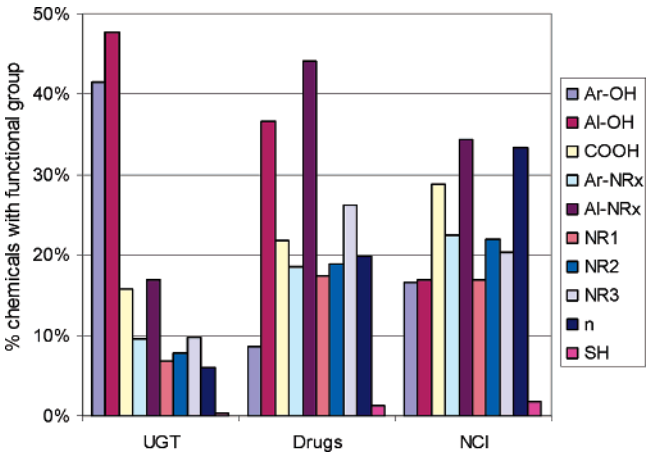

Figure 3. Comparison of functional group frequency in three different chemical libraries: chemicals tested for UGT metabolism (UGT), 3D Pharmaceutical Structure Database (Drugs), and diversity set of chemicals from the National Cancer Institute (NCI). The nucleophile types displayed include the following functional groups: phenolic (Ar−OH), aliphatic hydroxyl (Al−OH), carboxylic acid (COOH), amine adjacent to aromatic (Ar−NR_x), aliphatic amine (Al−NR_x), primary amine (NR_1), secondary amine (NR_2), tertiary amine (NR_3), aromatic nitrogen (n), and thiol (SH).

chemical diversity. It appears that chemicals containing an aromatic hydroxyl group, and to a lesser extent an aliphatic hydroxyl group, are over-represented in the UGT chemical set, compared to the drug and NCI chemical sets. Conversely, the occurrence of chemicals containing a carboxylic acid or an amine group is much higher (often more than 2-fold) in the drug set and NCI chemical set, compared to that in the UGT chemical set. This chemical diversity bias in the chemicals tested for UGT activity may result in models capable of predicting the glucuronidation of chemicals containing aromatic and aliphatic hydroxyls well, but with other types of nucleophilic atoms predicted suboptimally. This work represents the first use of chemical diversity concepts in relation to UGT screening and assumes significance for the selection of chemicals in future substrate selectivity studies. Efforts to screen chemicals with a greater and more relevant distribution of functional group types will be rewarded with the potential for robust and nonbiased models well-suited to the screening of large, structurally diverse data sets.

The results detailed here may be applied in a straightforward manner using only the structure of the chemical of

interest and the color map (Figure 2). The color map (or the same table with percentages replacing the colors) may be used to estimate the relative likelihood of each isoform glucuronidating nucleophilic atoms in a chemical. This is a first step toward the selection of substructures for chemical modification in order to increase or decrease the likelihood of glucuronidation, either in general or by a specific UGT isoform. Such "metabolic tailoring" of chemicals would represent a significant milestone in the progress toward managing and understanding xenobiotic detoxification in humans, and in designing drugs that are preferentially glucuronidated.

In summary, these results demonstrate that the nucleophilic atom and its local molecular environment are very important predictors of the site of glucuronidation. For many UGT isoforms, the attachment of an aromatic ring to the nucleophile and the type of nucleophilic atom itself were shown to have a highly statistically significant effect on the likelihood of glucuronidation. For most UGT isoforms, a simple representation of the local chemical structure and electronic environment of the nucleophilic atom was found to be the major predictor of the site of glucuronidation. The results presented here strongly indicate that general chemical properties, such as chemical size, shape, flexibility, and hydrophobicity, may be of secondary importance to the steric and electronic character immediately local to the nucleophilic atom. Molecular recognition based on the local structure is likely to be important for the ability of UGT to detoxify chemicals that differ greatly in gross molecular structure. The accrual of a larger, more diverse, and quantitative data set on the site of chemical glucuronidation will be vital for gaining further insight into how the nucleophilic atom local structure influences the glucuronidation capacity.

ACKNOWLEDGMENT

This work was funded by a grant from the National Health and Medical Research Council of Australia.

Supporting Information Available: Included are the data sets for UGT1A1, UGT1A3, UGT1A4, UGT1A6, UGT1A8, UGT1A9, UGT1A10, and UGT2B7 and the references from which these data were collected. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Armstrong, R. N. Enzyme-Catalyzed Detoxication Reactions: Mechanisms and Stereochemistry. *Crit. Rev. Biochem.* **1987**, *22*, 39−88.

(2) Radominska-Pandya, A.; Czernik, P. J.; Little, J. M.; Battaglia, E.; Mackenzie, P. I. Structural and Functional Studies of UDP-Glucuronosyltransferases. *Drug Metab. Rev.* **1999**, *31*, 817−99.

(3) Miners, J. O.; Smith, P. A.; Sorich, M. J.; McKinnon, R. A.; Mackenzie, P. I. Predicting Human Drug Glucuronidation Parameters: Application of in Vitro and in Silico Modeling Approaches. *Annu. Rev. Pharmacol. Toxicol.* **2004**, *44*, 1−25.

(4) Miners, J. O.; Mackenzie, P. I. Drug Glucuronidation in Humans. *Pharmacol. Ther.* **1991**, *51*, 347−69.

(5) Aumont, V.; Krisa, S.; Battaglia, E.; Netter, P.; Richard, T.; Merillon, J. M.; Magdalou, J.; Sabolovic, N. Regioselective and Stereospecific Glucuronidation of trans- and cis-Resveratrol in Human. *Arch. Biochem. Biophys.* **2001**, *393*, 281−289.

(6) Barbier, O.; Albert, C.; Martineau, I.; Vallee, M.; High, K.; Labrie, F.; Hum, D. W.; Labrie, C.; Belanger, A. Glucuronidation of the Nonsteroidal Antiestrogen EM-652 (SCH 57068), by Human and Monkey Steroid Conjugating UDP-Glucuronosyltransferase Enzymes. *Mol. Pharmacol.* **2001**, *59*, 636−645.

(7) Burchell, B.; Brierley, C. H.; Rance, D. Specificity of Human UDP-Glucuronosyltransferases and Xenobiotic Glucuronidation. *Life Sci.* **1995**, *57*, 1819−31.

(8) Doerge, D. R.; Chang, H. C.; Churchwell, M. I.; Holder, C. L. Analysis of Soy Isoflavone Conjugation in Vitro and in Human Blood Using Liquid Chromatography−Mass Spectrometry. *Drug Metab. Dispos.* **2000**, *28*, 298−307.

(9) Jude, A. R.; Little, J. M.; Bull, A. W.; Podgorski, I.; Radominska-Pandya, A. 13-Hydroxy- and 13-Oxooctadecadienoic Acids: Novel Substrates for Human UDP-Glucuronosyltransferases. *Drug Metab. Dispos.* **2001**, *29*, 652−5.

(10) Kemp, D. C.; Fan, P. W.; Stevens, J. C. Characterization of Raloxifene Glucuronidation in Vitro: Contribution of Intestinal Metabolism to Presystemic Clearance. *Drug Metab. Dispos.* **2002**, *30*, 694−700.

(11) Stone, A. N.; Mackenzie, P. I.; Galetin, A.; Houston, J. B.; Miners, J. O. Isoform Selectivity and Kinetics of Morphine 3- and 6-Glucuronidation by Human UDP-Glucuronosyltransferases: Evidence for Atypical Glucuronidation Kinetics by UGT2B7. *Drug Metab. Dispos.* **2003**, *31*, 1086−9.

(12) Uchaipichat, V.; Mackenzie, P. I.; Guo, X. H.; Gardner-Stephen, D.; Galetin, A.; Houston, J. B.; Miners, J. O. Human UDP-Glucuronosyltransferases: Isoform Selectivity and Kinetics of 4-Methylumbelliferone and 1-Naphthol Glucuronidation, Effects of Organic Solvents, and Inhibition by Diclofenac and Probenecid. *Drug Metab. Dispos.* **2004**, *32*, 413−23.

(13) Cupid, B. C.; Beddell, C. R.; Lindon, J. C.; Wilson, I. D.; Nicholson, J. K. Quantitative Structure−Metabolism Relationships for Substituted Benzoic Acids in the Rabbit: Prediction of Urinary Excretion of Glycine and Glucuronide Conjugates. *Xenobiotica* **1996**, 157−176.

(14) Cupid, B. C.; Holmes, E.; Wilson, I. D.; Lindon, J. C.; Nicholson, J. K. Quantitative Structure−Metabolism Relationships (QSMR) Using Computational Chemistry: Pattern Recognition Analysis and Statistical Prediction of Phase II Conjugation Reactions of Substituted Benzoic Acids in the Rat. *Xenobiotica* **1999**, *29*, 27−42.

(15) Ethell, B. T.; Ekins, S.; Wang, J. B.; Burchell, B. Quantitative Structure Activity Relationships for the Glucuronidation of Simple Phenols by Expressed Human UGT1A6 and UGT1A9. *Drug Metab. Dispos.* **2002**, *30*, 734−738.

(16) Said, M.; Ziegler, J. C.; Magdalou, J.; Elass, A.; Vergoten, G. Inhibition of Bilirubin UDP-Glucuronosyltransferase − A Comparative Molecular Field Analysis (COMFA). *Quant. Struct.-Act. Relat.* **1996**, *15*, 382−388.

(17) Smith, P. A.; Sorich, M. J.; Low, L. S. C.; McKinnon, R. A.; Miners, J. O. Modelling Metabolism by UDP-Glucuronosyltransferases. *J. Mol. Graphics Modell.* **2004**, *22*, 507−517.

(18) Smith, P. A.; Sorich, M. J.; McKinnon, R. A.; Miners, J. O. Pharmacophore and Quantitative Structure−Activity Relationship Modeling: Complementary Approaches for the Rationalization and Prediction of UDP-Glucuronosyltransferase 1A4 Substrate Selectivity. *J. Med. Chem.* **2003**, *46*, 1617−1626.

(19) Smith, P. A.; Sorich, M. J.; McKinnon, R. A.; Miners, J. O. In Silico Insights: Chemical and Structural Characteristics Associated with UDP-Glucuronosyltransferase (UGT) Substrate Selectivity. *Clin. Exp. Pharmacol. Physiol.* **2003**, *30*, 836−840.

(20) Sorich, M. J.; McKinnon, R. A.; Miners, J. O.; Winkler, D. A.; Smith, P. A. Rapid Prediction of Chemical Metabolism by Human UDP-Glucuronosyltransferase Isoforms Using Quantum-Chemical Descriptors Derived with the Electronegativity Equalization Method. *J. Med. Chem.* **2004**, *47*, 5311−7.

(21) Sorich, M. J.; Miners, J. O.; McKinnon, R. A.; Smith, P. A. Multiple Pharmacophores for the Investigation of Human UDP-Glucuronosyltransferase Isoform Substrate Selectivity. *Mol. Pharmacol.* **2004**, *65*, 301−308.

(22) Sorich, M. J.; Miners, J. O.; McKinnon, R. A.; Winkler, D.; Burden, F. R.; Smith, P. A. Comparison of Linear and Non-linear Classification Algorithms for the Prediction of Drug and Chemical Metabolism by Human UDP-Glucuronosyltransferase Isoforms. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2019−2024.

(23) Sorich, M. J.; Smith, P. A.; McKinnon, R. A.; Miners, J. O. Pharmacophore and Quantitative Structure Activity Relationship Modelling of UDP-Glucuronosyltransferase 1A1 (UGT1A1) Substrates. *Pharmacogenetics* **2002**, *12*, 635−645.

(24) de Hoon, M. J.; Imoto, S.; Nolan, J.; Miyano, S. Open Source Clustering Software. *Bioinformatics* **2004**, *20*, 1453−4.

(25) R Development Core Team. *R: A Language and Environment for Statistical Computing;* R Foundation for Statistical Computing: Vienna, Austria, 2005.

(26) Agresti, A. A. Survey of Exact Inference for Contingency Tables. *Stat. Sci.* **1992**, *7*, 131−153.

(27) Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J. P. The Electronegativity Equalization Method I: Parametrization and Validation for Atomic Charge Calculations. *J. Phys. Chem.* **2002**, *106*, 7887−7894.

(28) Bauer, D. F. Constructing Confidence Sets Using Rank Statistics. *J. Am. Stat. Assoc.* **1972**, *67*, 687−690.
(29) Hand, D. J.; Yu, K. Idiot's Bayes − Not So Stupid after All? *Int. Stat. Rev.* **2001**, *69*, 385−399.
(30) Demsar, J.; Zupan, B.; Leban, G. *Orange: From Experimental Machine Learning to Interactive Data Mining, White Paper*; Faculty of Computer and Information Science, University of Ljubljana: Ljubljana, Slovenia, 2004. www.ailab.si/orange (accessed Mar 2006).
(31) Hawkins, D. M.; Basak, S. C.; Mills, D. Assessing Model Fit by Cross-Validation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 579−86.
(32) Beck, J. R.; Schultz, E. K. The Use of ROC Curves in Test Performance Evaluation. *Arch. Pathol. Lab. Med.* **1986** *110*, 13−20.
(33) Voigt, J. H.; Bienfait, B.; Wang, S. M.; Nicklaus, M. C. Comparison of the NCI Open Database with Seven Large Chemical Structural Databases. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 702−712.
(34) Bone, R.; Agard, D. A. Mutational Remodeling of Enzyme Specificity. *Methods Enzymol.* **1991**, *202*, 643−71.
(35) Geerlings, P.; De Proft, F. Chemical Reactivity as Described by Quantum Chemical Methods. *Int. J. Mol. Sci.* **2002**, *3*, 276−309.
(36) Li, Y.; Evans, J. N. S. The Fukui Function: A Key Concept Linking Frontier Molecular Orbital Theory and the Hard−Soft−Acid−Base Principle. *J. Am. Chem. Soc.* **1995**, *117*, 7756−7759.
(37) Langenaeker, W.; Demel, K.; Geerlings, P. Quantum-Chemical Study of the Fukui Function as a Reactivity Index. Part 2: Electrophilic Substitution on Mono-substituted Benzenes. *J. Mol. Struct.* **1991**, *234*, 329.

CI600248E