# Pharmacophore and Quantitative Structure−Activity Relationship Modeling: Complementary Approaches for the Rationalization and Prediction of UDP-Glucuronosyltransferase 1A4 Substrate Selectivity

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

*Department of Clinical Pharmacology, Flinders University and Flinders Medical Centre, Bedford Park, 5042, South Australia, and School of Pharmaceutical, Molecular and Biomedical Sciences, University of South Australia, Adelaide, 5001, South Australia*

*Received September 11, 2002*

Pharmacophore, two-dimensional (2D), and three-dimensional (3D) quantitative structure−activity relationship (QSAR) modeling techniques were used to develop and test models capable of rationalizing and predicting human UDP-glucuronosyltransferase 1A4 (UGT1A4) substrate selectivity and binding affinity (as $K_{m,app}$). The dataset included 24 structurally diverse UGT1A4 substrates, with 18 of these comprising the training set and 6 an external prediction set. A common features pharmacophore was generated with the program Catalyst after overlapping the sites of conjugation using a novel, user-defined "glucuronidation" feature. Pharmacophore-based 3D-QSAR ($r^2 = 0.88$) and molecular-field-based 3D-QSAR ($r^2 = 0.73$) models were developed using Catalyst and self-organizing molecular field analysis (SOMFA) software, respectively. In addition, a 2D-QSAR ($r^2 = 0.80$, CV $r^2 = 0.73$) was generated using partial least-squares (PLS) regression and variable selection using an unsupervised forward selection (UFS) algorithm. Both UGT1A4 pharmacophores included two hydrophobic features and the glucuronidation site. The 2D-QSAR showed the best overall predictivity and highlighted the importance of hydrophobicity (as log *P*) in substrate−enzyme binding.

## Introduction

Recognition by the pharmaceutical industry that undesirable absorption, distribution, metabolism, and excretion (ADME) properties of new drug candidates are the cause of many clinical phase drug development failures has led to the need to identify such problems in the drug discovery process.[1] In this regard, the demand for the rapid evaluation of increasing numbers of compounds has driven the development of methodologies and software, which has resulted in the increased application of molecular modeling techniques to ADME.[2] Indeed, computational (in silico) and in vitro ADME approaches are now widely utilized throughout the discovery process to optimize the selection of the most suitable drug candidates for development. Notably, in silico approaches potentially allow the ADME properties of a newly discovered compound to be predicted by modeling the structural and physicochemical characteristics of that compound without recourse to laboratory-based procedures.

An integral component of ADME evaluation is the identification of the enzyme(s) responsible for the metabolism of a new chemical entity, together with an estimate of kinetic parameters. Cytochrome P450 (CYP) and UDP-glucuronosyltransferase (UGT) are the principal drug metabolizing enzyme systems.[3,4] Apart from metabolizing drugs, however, CYP and UGT also catalyze the biotransformation of a multitude of structurally diverse nondrug xenobiotics (e.g., environmental pollutants, dietary chemicals) and endogenous compounds (e.g., steroid hormones). Consistent with their broad substrate profiles, CYP and UGT are known to exist as superfamilies of independently regulated enzyme forms ("isoforms") that exhibit distinct, but often overlapping, substrate and inhibitor selectivities. Thus, prediction of variability in drug elimination and response in humans requires knowledge of CYP and UGT isoform substrate and inhibitor selectivities and the factors that regulate isoform catalytic activity, particularly drug−drug interactions and genetic polymorphisms.

In recent years, considerable effort has been directed toward predicting the substrate and inhibitor selectivities of human CYP isoforms using molecular modeling techniques including pharmacophore, 2D, and 3D quantitative structure−activity relationships (QSAR) and homology modeling.[5,6] For example, pharmacophore and 3D-QSAR models have been developed that can be used to infer the active site binding requirements of substrates and inhibitors for numerous CYP isoforms and that can predict apparent $K_m$ ($K_{m,app}$) for substrates and apparent $K_i$ ($K_{i,app}$) for inhibitors.[7] In particular, Ekins and colleagues have used the molecular modeling software Catalyst (Accelrys, Inc.) to derive pharmacophore models for several CYP isoforms that predict $K_{m,app}$ or $K_{i,app}$ values within 1 log unit.[8−12] However, the pharmacophores do not necessarily overlay the oxidation sites of the chemicals, which may be a significant drawback for substrate-derived pharmacophores because it precludes any chemically intuitive interpretation of the alignments. Other groups[13,14] have combined pharmacophore models with homology models based on crystal structures of bacterial CYPs and, more recently, the mammalian enzyme CYP2C5. These combined models provide powerful tools that take into

* To whom correspondence should be addressed. Phone: 61 8 8204 3155. Fax: 61 8 8204 5114. E-mail: paul.smith@flinders.edu.au.
† Flinders University and Flinders Medical Centre.
‡ University of South Australia.

10.1021/jm020397c CCC: $25.00 
Published on Web 04/02/2003

account the site of oxidation and allow investigation of active site features and substrate interactions.

The successful adoption of in silico approaches for study of the CYP family has progressed in parallel with the increasing availability of CYP isoform substrate and inhibitor selectivities, and the development of such models has also been aided as a result of access to homology models of human CYPs. However, the development of models for predicting metabolism and for characterizing structural features of substrates for UGT isoforms is less advanced relative to CYP. Only recently has the substrate profile of UGT isoforms begun to approach an interpretable level, and an X-ray crystal structure is not yet available. Nevertheless, it is envisaged that molecular modeling techniques, in particular 2D- and 3D-QSAR and pharmacophore modeling, will translate well for application to UGT.

Recent studies have investigated the substrate profile of the human UGT isoform UGT1A4.[15,16] It has been demonstrated that UGT1A4 has the ability to metabolize a structurally diverse group of compounds, which includes drugs, nondrug xenobiotics, and hydroxysteroids. Of particular interest is the capacity of UGT1A4 to catalyze the glucuronidation of primary (e.g., 4-aminobiphenyl), secondary (e.g., desmethylclozapine), and tertiary amine (e.g., clozapine) containing compounds. $K_{m,app}$ values determined for the UGT1A4 substrates ranged from 7 to 11 600 $\mu$M.

The studies reported here aimed to investigate the structural and chemical properties that characterize substrates of UGT1A4 and to evaluate several pattern recognition techniques for the prediction of $K_{m,app}$ for a set of UGT1A4 substrates. Catalyst was used to generate two pharmacophores for UGT1A4 substrates in which the sites of conjugation were overlaid. One pharmacophore ("common features") can be used to align molecules sensibly, and the other allows quantitative prediction of the substrate $K_{m,app}$. The alignment from the common features model was used to develop a molecular-field-based (SOMFA) model. Furthermore, a 2D-QSAR model was developed using the software packages Cerius2 (Accelrys, Inc.) and Dragon (Milano Chemometrics and QSAR Research Group, Milan, Italy) based on a range of descriptors that capture information associated with ligand−receptor binding events. The predictivity of the three QSAR models was comparable to those reported previously in the drug metabolism field. Overall, these models provide a complementary package for inferring various features associated with substrate binding to UGT1A4. The development of multiple models has also allowed comparison of these different computational approaches for the challenge of modeling UGTs.

## Materials and Methods

**Dataset Selection.** The dataset comprised UGT1A4 substrates characterized by Green and colleagues.[15,16] These data were derived from a single laboratory, using a common assay procedure to quantitate aglycone glucuronidation by UGT1A4 stably expressed in HK293 cells. Hence, the kinetic parameter modeled, namely, $K_{m,app}$, may be considered to be internally consistent. Substrates modeled were structurally diverse and had $K_{m,app}$ values that spanned over 3 orders of magnitude, both prerequisites for the Catalyst Hypogen algorithm. The dataset comprised 24 substrates (Figure 1) and their corresponding $K_{m,app}$ values, which ranged from 7 to 11 600 $\mu$M.

**Model Generation.** The molecular modeling studies were performed using Silicon Graphics Octane2 (Silicon Graphics, Mountain View, CA) and x86 Intel workstations. To test the predictivity of the models constructed (described below), the 24 substrates investigated were divided into two groups: one to generate the models (18 chemicals) and the second to test the models (6 chemicals). The test set of six chemicals was selected to span a wide range of $K_{m,app}$ values (47−1570 $\mu$M) and for diversity of chemical structure in order to stringently test the models generated. The ability of the models to predict the $K_{m,app}$ values of these six chemicals was used as the common method of validation. Currently, prediction within 1 log unit of the experimental $K_{m,app}$ value is considered standard in the drug metabolism field.[10] Randomization tests were employed to assess statistical significance whenever supported by the software. This involved the generation of alternative training sets by randomization of the association between $K_{m,app}$ and the substrate such that a chemical was not correctly associated with its $K_{m,app}$ value. Repetition of the model generation procedure with these randomized data sets allowed calculation of the probability that the model was a result of a chance correlation. To achieve a 99% confidence level, 99 randomization tests were necessary while 19 randomization tests were required to achieve a 95% confidence level.

**Pharmacophore Modeling with Catalyst.** Catalyst, version 4.6 (Accelrys, Inc.), was used for modeling. Substrates for UGT1A4 were drawn from the View Compound Workbench using the Visualizer module. Conformer generation was performed using the "Best" function to ensure the widest possible conformer coverage. Each substrate was allowed a maximum of 255 conformers within an energy range of 20 kcal/mol. The default method used by Catalyst for fitting a substrate to a pharmacophore is the "Fast Fit" method. "Fast Fit" finds the optimum fit of a conformer to a pharmacophore without performing an energy minimization on the conformer. The "Best Fit" procedure starts with "Fast Fit" and performs a more thorough search of the torsional and rotational space of the conformer. Common features pharmacophores were developed using the Catalyst HipHop module, and predictive 3D-QSAR pharmacophores were developed using the Catalyst HypoGen module. Catalyst refers to these hypothetical pharmacophores as "hypotheses", and these two terms will subsequently be used interchangeably. Interfeature spacing for both modules of Catalyst was set to 1 Å because all molecules were relatively small and feature-poor.

HipHop hypotheses were developed in an iterative manner by gradually allowing inclusion of a greater number and diversity of chemical features. The final "common features" hypotheses were then analyzed for uniformity and sensibility of mappings of all compounds. Here, the principal measure of "sensibility" is the mapping of the glucuronidation site. In most cases, the best hypothesis was also found to be the lowest "cost" hypothesis. The total energy cost of a generated pharmacophore can be calculated from the deviation between the estimated activity and the observed activity, combined with the complexity of the hypothesis. A null hypothesis can also be calculated on the basis of the presumption that the experimental activities are normally distributed about their mean and that there is no relationship between the fit and the independent data.

Hypogen is a Catalyst module that searches for a 3D arrangement of abstract chemical features that explain trends and variations of activity with the chemical structure. It can be used to predict the binding ability of chemicals toward an enzyme or receptor. The fit of the chemicals to the putative pharmacophore (i.e., how closely the features of each chemical can match the features of the pharmacophore) is then linearly correlated to the activity of the chemicals.[17]

**Generation of User-Defined Glucuronidation Features within Catalyst.** Early experiments in pharmacophore development using the UGT1A4 dataset highlighted some difficulties caused by using the default hydrogen bond acceptor (HBA) and hydrogen bond donor (HBD) features to recognize glucuronidation sites within the substrates. For example, a problem was discovered with the choice of the default Catalyst

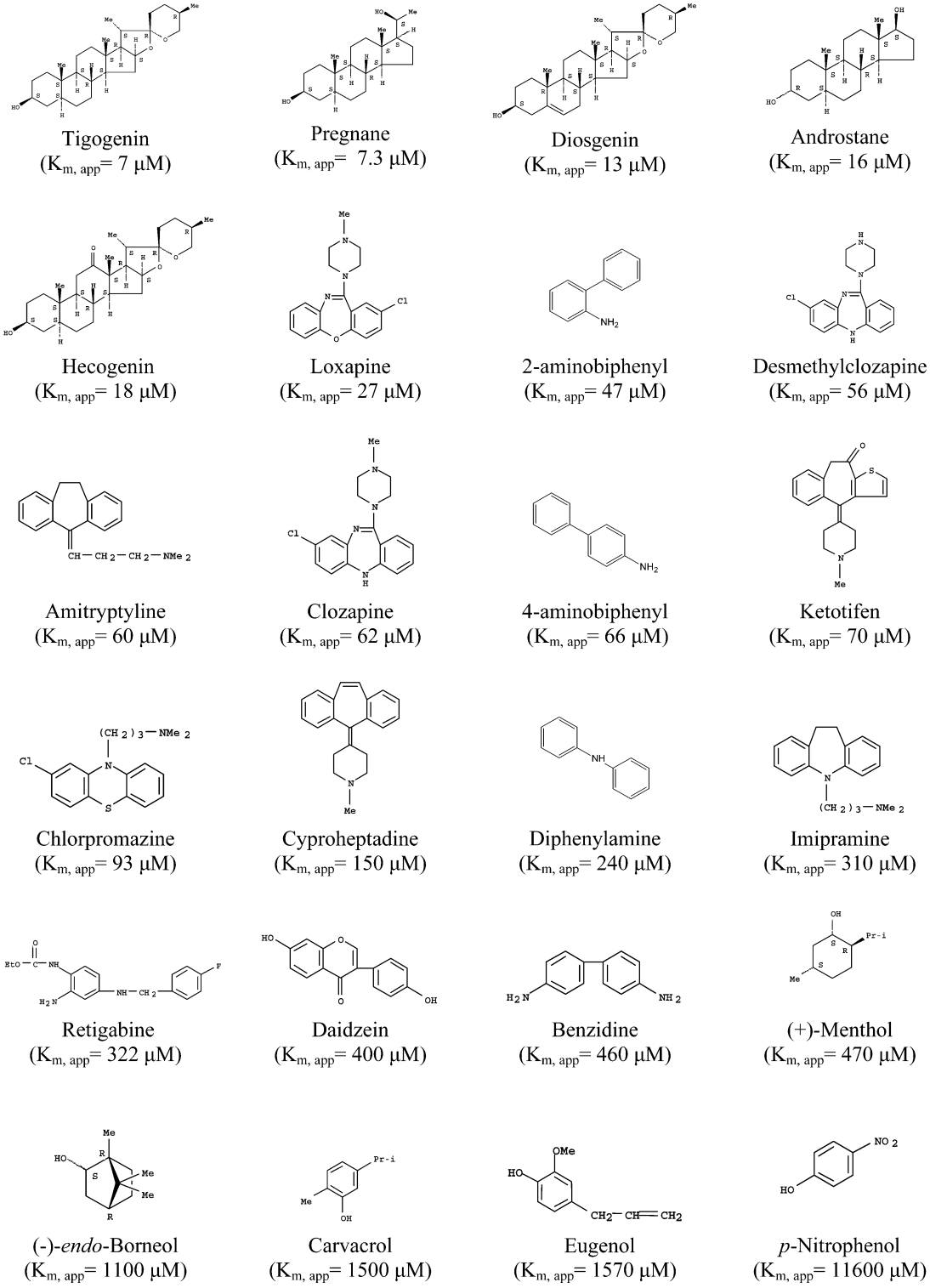


**Figure 1.** Structures and $K_{m,app}$ values of the 24 UGT1A4 substrates in the training and test sets.

HBA feature because Catalyst does not consider $sp^2$ nitrogen atoms adjacent to aromatic rings (e.g., aniline-type nitrogens) as possessing a lone pair of electrons. Such atomic features are automatically assigned within the Catalyst drawing module and are the default, and consequently, it was not possible to map aniline-type nitrogen atoms to an HBA-based feature. Thus, design of an HBD-based glucuronidation feature was undertaken. Such a feature maps a vector in the direction of the hydrogen. A difficulty in using an HBD-based glucuronidation feature is to ensure that the feature recognizes tertiary amines as glucuronidation sites because they do not have a hydrogen. This was overcome by allowing the vector to map in the direction of the lone pair in the specific case of a tertiary nitrogen atom. This compromise is necessary in order to achieve sufficiently general recognition of conjugation sites. Such a compromise needs to be borne in mind when considering the mechanistic interpretation of mappings of the "exception" substrates. It should also be noted that an analogous approach could be applied to the previous HBA-based feature; the aniline type hydrogens could be allowed to map in the direction of the lone pair vector in specific cases.

A third alternative can be considered whereby the glucuronidation feature has no associated vector. Pharmacophores containing such a feature may sacrifice specificity (because

only one feature, not two, are used), but the feature solves the difficulties alluded to above regarding substrate site recognition. This "general" glucuronidation feature performs very well and is widely recognized. All new features were added to the Feature Dictionary of Catalyst.

**HBA-Based Glucuronidation Vector (HBA-Glc. Vct.).** An "HBA lipid" chemical function was edited to recognize only glucuronidation sites using the "View Hypothesis" workbench. This involved removal of the capability to recognize features such as ethers, esters, aromatic heteroatoms, etc. However, this function cannot recognize $sp^2$ nitrogen atoms adjacent to aromatic rings (i.e., aniline-type nitrogens) because Catalyst automatically assigns such atoms as lacking a lone pair of electrons ($n = 0$). Hence, they are considered unable to act as HBAs, despite the well-known nucleophilic nature of such atoms.

**HBD-Based Glucuronidation Vector (HBD-Glc. Vct.).** In addition, an HBD chemical function was edited to recognize tertiary amines and map the vector in the direction of the lone pair electrons from the "View Hypothesis" workbench. This allows the chemical function to recognize all glucuronidation sites in the dataset, although a mechanistic interpretation may be less valid. Editing was achieved by altering the atom specifications for an "HBA lipid" chemical function to recognize only a tertiary nitrogen, and then this function was included as an "Or" option within the original HBD definition. Furthermore, the original HBD chemical function was modified to exclude primary and secondary amides as HBDs because these are not glucuronidation sites. Validation of the feature was achieved by testing its ability to recognize known glucuronidation sites of UGT1A4 substrates.

**Glucuronidation Sphere (Glc. Sph.).** An "HBD" chemical function was modified in order to recognize the heavy atom (O, S, N) of any potential glucuronidation site. By use of the "View Hypothesis" workbench of Catalyst, vector associations were removed from the original HBD feature. Atom specifications were altered to prevent recognition of aromatic heteroatoms (O, S, N) and imine and amide nitrogens because these chemical features are not sites of glucuronidation (with the rare exception of some aromatic N centers, which are not present in this data set). Again, testing its ability to recognize known glucuronidation sites of UGT1A4 substrates validated the feature.

**Modeling with SOMFA.** Self-organizing molecular field analysis (SOMFA) is based on a 3D-QSAR algorithm, similar to CoMFA, that can detect the 3D positions in the electrostatic and shape fields that influence the activities of a set of aligned molecules.[18] The algorithm calculates the steric and electrostatic fields for a set of aligned conformers and allows visualization of significant interaction points on a ligand. The common features (HipHop) pharmacophore alignment was used to align the molecules for SOMFA model generation. The aligned chemicals were then input to the program along with the $-\log K_{m,app}$. The shape and electrostatic field were sampled at 1 Å intervals over a 3D grid. The output consisted of a set of points in 3D that represent the regions of space where the shape or electrostatics of the substrate influenced the $K_{m,app}$.

**2D-QSAR Modeling.** A wide range of 2D descriptors was calculated from the structure of each chemical using the Dragon (Milano Chemometrics and QSAR Research Group, Milan, Italy) and Cerius2 (Accelrys, Inc.) programs. These included topological, 2D autocorrelation, constitutional, BCUT, thermodynamic, and electronic descriptors.[34] Before generation of the model, the least collinear and most relevant subsets of descriptors were selected and all descriptors without a statistically significant correlation to $-\log K_{m,app}$ were removed. The significantly correlated descriptors were input to a program implementing the Unsupervised Forward Selection (UFS) algorithm.[19] This method selects the two least well-correlated descriptors and then selects additional variables on the basis of their multiple correlation with those already chosen, thereby selecting a subset of variables that are as close to orthogonal as possible. Subsets of descriptors were selected using this procedure with varying degrees of collinearity allowed ($r^2_{max} = 0.1, 0.2, ..., 0.9, 0.99$).

A regression model was built for each descriptor subset using principal component (PCR) regression. The variable subset giving the best leave-one-out (LOO) cross-validated $r^2$, a measure of predictive ability of the model, was chosen for further optimization. Any descriptor found to have a significant correlation with the residuals of the model was included if it improved the LOO cross-validated $r^2$. Descriptors were omitted from the model if the removal resulted in an increase in the LOO cross-validated $r^2$.

## Results

**Pharmacophore Modeling of UGT1A4 Substrates.** Catalyst identifies the 3D spatial orientation and overall range of structural features associated with the ligands and uses them to define a pharmacophore. Subsequently, Catalyst can be used in two ways. Identification of the features common to all substrates generates a qualitative "common features" model that highlights the minimum necessary 3D arrangement of features that are necessary for binding. Alternatively, the "fit" of a ligand to a pharmacophore can be correlated with a measure of binding affinity (e.g., $K_{m,app}$) to generate a quantitative model potentially capable of predicting that measure of binding affinity.

**"Common Features" Pharmacophore Model.** The Catalyst HipHop module was used to generate a "common features" model. The initial aim was to deduce the distance(s) between one glucuronidation site and one other common feature for all 24 substrates, since this represents the simplest pharmacophore that can be modified progressively. To achieve this, the vector-free glucuronidation feature (Glc. Sph., see Materials and Methods), which recognized all 24 substrates, was used. The occurrence of this feature in a pharmacophore was set to a minimum of 1 (i.e., each pharmacophore had to include one glucuronidation feature), and Catalyst was given the option of choosing only one other feature from the default hydrophobe, HBA "lipid" or HBD features. The two solutions consisted of a glucuronidation feature and a hydrophobic feature separated by either 3 or 6.2 Å. With the exception of 4-nitrophenol, all substrates could be mapped to the glucuronidation feature and the hydrophobe feature of each pharmacophore. Catalyst interpreted 4-nitrophenol as containing a glucuronidation feature but no hydrophobe; hence, the substrate was mapped only to the glucuronidation feature of each pharmacophore.

No single vectorized glucuronidation function could be designed to recognize the glucuronidation site of every molecule. For example, the HBA-based glucuronidation feature could not recognize the sites in the biphenylamines; hence, the HBD-based feature must be used for these substrates even though tertiary amines are not recognized by the HBD-based feature. Hence, the dataset was arbitrarily classified into groups comprising biphenyls, tertiary amines, steroids, and small molecules (see Table 2). This allowed combinations of groups to be investigated with vectorized glucuronidation features.

Each group was searched for common features, using a vectorized glucuronidation feature, and then combinations of the groups were searched. However, the small molecules group, which ranged in $K_{m,app}$ from 470 to 11 600 $\mu$M, was excluded from pharmacophore genera-

**Table 1.** Test Set Predicted $K_{m,app}$ Values and Logarithm of the Residuals for the Three Models

| substrate | obsd $K_{m,app}$ | pharmacophore predicted $K_{m,app}$ [a] | SOMFA predicted $K_{m,app}$ [a] | 2D-QSAR predicted $K_{m,app}$ [a] |
|---|---|---|---|---|
| 2-aminobiphenyl | 47 | 190 (0.6) | 190 (0.6) | 270 (0.7) |
| clozapine | 62 | 80 (0.1) | 160 (0.4) | 100 (0.2) |
| retigabine | 322 | 70 (0.6) | 250 (0.1) | 260 (0.1) |
| menthol | 470 | 150 (0.5) | 260 (0.4) | 1200 (0.4) |
| carvacrol | 1500 | 190 (0.9) | 190 (0.9) | 430 (0.5) |
| eugenol | 1570 | 150 (1.0) | 390 (0.6) | 1100 (0.2) |

[a] $K_{m,app}$ values in units of $\mu$M. Values in parentheses represent the log of the residuals (i.e., log of observed value minus predicted value).

tion procedures because of the molecules' small size and feature poorness. Analysis of the biphenyl group for the common spatial arrangement of features generated only one pharmacophore consisting of two hydrophobic regions and one HBD-based glucuronidation feature (HBD-Glc. Vct). The arrangement of features is a triangular combination of the two distances found to be common to all 24 substrates in the initial simplified model (i.e., 3 and 6.2 Å) and is shown in Figure 2a.

The common spatial arrangement of features for the combined biphenyl and steroid group was then investigated. The highest ranked pharmacophore for this substrate group shared the same features in the same geometry as that shown in Figure 2a (i.e., two hydrophobes and a HBD-Glc. Vct.). Substrates in this group varied in $K_{m,app}$ from 7 to 460 $\mu$M. Similarly, the common spatial arrangements of features for the combined biphenyl and tertiary amine groups were examined, and the highest ranked pharmacophore for this substrate group was also the same as that shown in Figure 2a. Substrates in this group varied in $K_{m,app}$ between 27 and 460 $\mu$M. Parts b and c of Figure 2 illustrate two representative UGT1A4 substrates, eugenol (which fitted despite exclusion from the training set) and clozapine, mapped to the "common features" pharmacophore. Table 3 shows the number of features of the common features pharmacophore mapped by a given substrate, conformer number and energy, reported default fit value, and maximium omitted features (MOF) setting used for the hypothesis generation.

**Pharmacophore-Based 3D-QSAR Model.** The HypoGen module of Catalyst was used to develop a 3D-QSAR for UGT1A4 using published $K_{m,app}$ values shown in Figure 1 for a training set of 18 structurally diverse substrates of UGT1A4.[15,16] For model generation, the HBD-based glucuronidation feature (HBD-Glc. Vct.) developed here and the default hydrophobic, hydrogen bond acceptor, hydrogen bond donor, and ring aromatic features were chosen as feature options. Of the 10 hypotheses returned, the lowest cost model was assessed as the best. This assessment was based not only on the

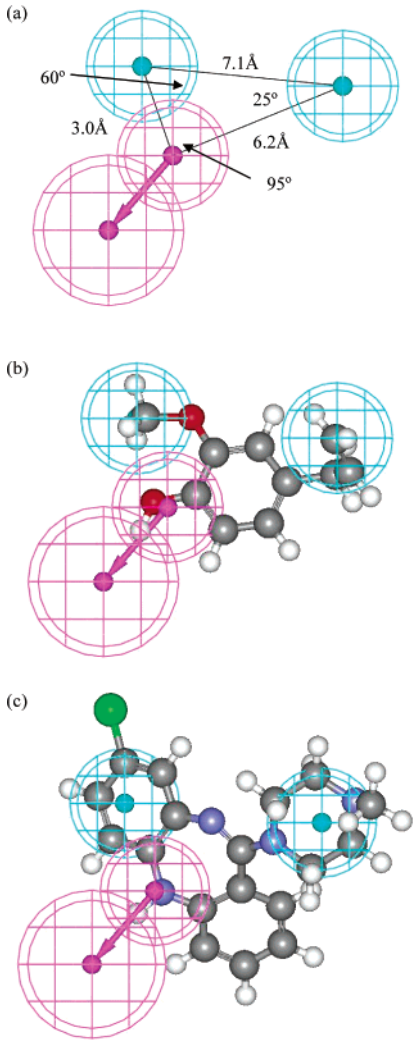


**Figure 2.** (a) The "common features" pharmacophore model developed using the HipHop module of Catalyst. Cyan spheres represent hydrophobic regions, and purple represents the "glucuronidation" feature. (b) Eugenol mapped to the Hypogen pharmacophore. (c) Clozapine mapped to the Hypogen pharmacophore.

Catalyst cost analysis but also on the possession of features that were representative of the remaining hypotheses, the highest correlation coefficient ($r$), and the best prediction of the test set. A high correlation ($r = 0.94$; $r^2 = 0.88$) between the experimentally derived and predicted $K_{m,app}$ values for the training set was obtained, suggestive of a high-quality model. Figure 3a demonstrates the structural features and geometry of the pharmacophore derived using Hypogen. Mappings of imipramine and benzidine to the hypogen pharmacophore are shown in parts b and c of Figure 3, respectively.

**Table 2.** Substrate Groups from Which the "Common Features" Pharmacophore Was Developed

| biphenyls | tertiary amines | steroids | small molecules |
|---|---|---|---|
| 2-aminobiphenyl | loxapine | 5α-pregnane-3$\beta$,20α-diol | menthol |
| 4-aminobiphenyl | desmethylclozapine | tigogenin | borneol |
| diphenylamine | amitriptyline | diosgenin | carvacrol |
| retigabine | clozapine | 5α-androstane-3α,17$\beta$-diol | eugenol |
| daidzein | ketotifen | hecogenin | 4-nitrophenol |
| benzidine | chlorpromazine | | |
| | cyproheptadine | | |
| | imipramine | | |

**Table 3.** Number of Features of the Common Features Pharmacophore Mapped by a Given Substrate, Conformer Number and Energy, Reported Default Fit Value, and Maximium Omitted Features (MOF) Setting Used for the Hypothesis Generation

| UGT1A4 substrate | features mapped | conformer number | conformer energy[a] | fit value | MOF[b] |
|---|---|---|---|---|---|
| tigogenin | 3 | 9 | 17.6 | 1.9 | −1 |
| pregnane | 3 | 8 | 19.7 | 2 | −1 |
| diosgenin | 3 | 4 | 18.3 | 1.7 | −1 |
| androstane | 3 | 7 | 14.2 | 1.7 | −1 |
| hecogenin | 3 | 9 | 13.2 | 1.8 | −1 |
| loxapine | 3 | 2 | 17 | 1.4 | 0 |
| 2-aminobiphenyl | 3 | 6 | 12.6 | 2.2 | −1 |
| desmethylclozapine | 3 | 9 | 19.3 | 2.7 | −1 |
| amitriptyline | 3 | 25 | 5.9 | 0.4 | 0 |
| clozapine | 3 | 19 | 3.7 | 2.9 | −1 |
| 4-aminobiphenyl | 2 | 2 | 0 | 1.6 | −1 |
| ketotifen | 3 | 16 | 9.5 | 0.5 | 0 |
| chlorpromazine | 3 | 8 | 18.7 | 2.1 | −1 |
| cyproheptadine | 3 | 9 | 4.7 | 2 | −1 |
| diphenylamine | 3 | 5 | 6.8 | 0.6 | 0 |
| imipramine | 3 | 37 | 4.6 | 1.8 | 0 |
| retigibine | 3 | 52 | 16 | 3 | −1 |
| daidzein | 3 | 12 | 0 | 0.4 | 0 |
| benzidine | 3 | 2 | 0 | 0.4 | 0 |
| (+)-menthol | 3 | 11 | 0 | 2.3 | −1 |
| (−)-*endo*-borneol | 2 | 5 | 0 | 2 | −1 |
| carvacrol | 3 | 7 | 0.2 | 1.9 | 0 |
| eugenol | 3 | 5 | 1.3 | 2.8 | 0 |
| 4-nitrophenol | 1 | 0 | 0 | 1 | −1 |

[a] Conformer energy has units of kcal/mol. [b] Maximum omitted features (MOF) = −1 means that Catalyst considers all possible mappings and finds the best possible fit. MOF = 0 means that Catalyst ignores mappings that involve a subset of features.

Further validation of the predictive 3D-QSAR model was undertaken by permutation and test set prediction. The statistical significance of the retrieved hypothesis was verified by randomizing the assignment of a dependent variable to a structure (i.e., the $K_{m,app}$ of each structure was randomly assigned to a different structure and the hypothesis generation procedure was repeated). This permutation process was repeated 19 times without retrieving a more "significant" (lower cost) hypothesis, indicating that the original model was unlikely to be a chance event ($p < 0.05$). A permuted hypothesis is considered more "significant" if the associated "hypothesis cost" is closer to the "fixed cost" than the original hypothesis cost (see Materials and Methods for description of "cost").

$K_{m,app}$ values for four of the six compounds comprising the validation set were estimated within 0.5 log unit of the experimentally derived $K_{m,app}$ (see Table 1), with eugenol having an error slightly greater than 1 log unit.

**SOMFA Model.** A 3D-QSAR model was developed using the SOMFA methodology of Robinson and coworkers.[18] Using the HipHop derived pharmacophore alignment, the same training set molecules (Figure 1) were imported into the SOMFA software program, and the electrostatic and steric maps were calculated for each. The resulting model showed a good correlation ($r = 0.85$; $r^2 = 0.73$) for the training set using the shape grid potentials. All six of the molecules in the test set were then predicted within 1 log unit using the model (See Table 1).

**2D-QSAR Model.** A 2D QSAR model was constructed with Cerius2 software (Accelrys, Inc.) using a range of chemical descriptors and the principal component (PCR) regression methodology. Of the 319 descriptors calculated for each chemical using the Dragon and Cerius2 software, 208 were found to be significantly correlated to $-\log K_{m,app}$ at the 95% confidence level. The UFS with an $r^2_{max}$ of 0.99 gave a 21-descriptor model with the best LOO cross-validated $r^2$ of 0.68. Sixteen descriptors in the model were found to increase the LOO cross-validated $r^2$ when removed from the model generation process. Thus, the optimum model was found to be a 5-descriptor, 1-component model with $r^2 = 0.80$ ($r = 0.89$) and LOO cross-validated $r^2 = 0.73$.

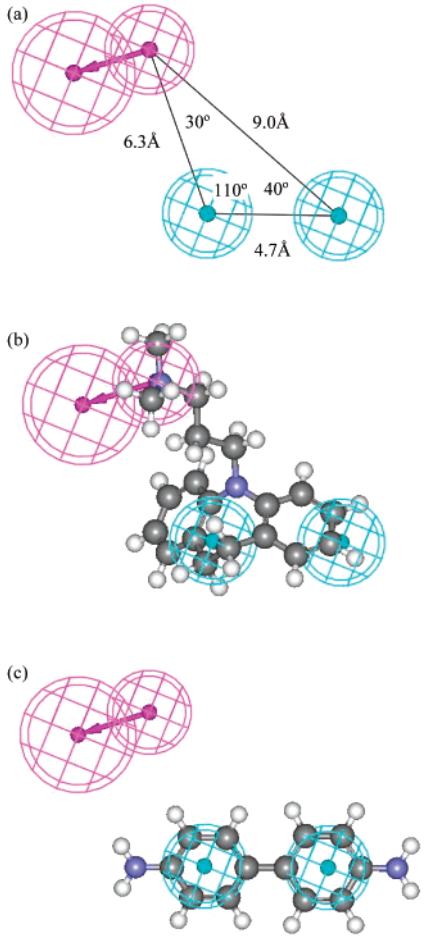


**Figure 3.** (a) Pharmacophore-based 3D-QSAR model developed using the Hypogen module of Catalyst. Cyan spheres represent hydrophobic regions, and purple represents the "glucuronidation" feature. (b) Imipramine mapped to the Hypogen pharmacophore. (c) Benzidine mapped to the Hypogen pharmacophore.

The model using the nonstandardized descriptors (i.e., calculated directly) is given by

$$pK_{m,app} = -3.54 + (0.21)(\text{AlogP98}) + (0.18)(\text{nRO5}) + (3.40)(\text{BENV1}) + (2.51)(\text{MATS3v}) + (1.66)(\text{MATS5e})$$

Standardization of the descriptors for the particular training set involves mean centering and unit normalization (by subtracting the mean and dividing by the standard deviation). Using standardized descriptors gives an indication of the relative contribution of each descriptor to the equation

$$pK_{m,app} = 3.88 + (0.24)(\text{AlogP98}_{normal}) + (0.14)(\text{nRO5}_{normal}) + (0.20)(\text{BENV1}_{normal}) + (0.22)(\text{MATS3v}_{normal}) + (0.19)(\text{MATS5e}_{normal})$$

**Table 4.** Average "Leave *N* Out" ($L_N$O) Cross-Validated $r^2$ of 100 Trials, with *N* Substrates Omitted ($N = 2-7$) from the 18 Training Set Chemicals, and the Associated Standard Deviations

| $L_N$O | % chemicals left out | cross-validated $r^2$ | standard deviation |
|---|---|---|---|
| $L_1$O | | 0.791 | |
| $L_2$O | 11 | 0.800 | 0.015 |
| $L_3$O | 17 | 0.795 | 0.025 |
| $L_4$O | 22 | 0.793 | 0.026 |
| $L_5$O | 28 | 0.784 | 0.061 |
| $L_6$O | 33 | 0.775 | 0.05 |
| $L_7$O | 39 | 0.752 | 0.08 |

where

(1) AlogP98 is log of the octanol/water partition coefficient calculated by an atom-based method,

(2) MATS3v is the Moran autocorrelation of path length 3 weighted by atomic van der Waals volumes,

(3) BENV1 is the negative burden eigenvalue 1 weighted by atomic van der Waals volumes,

(4) MATS5e is the Moran autocorrelation of path length 5 weighted by atomic Sanderson electronegativities, and

(5) nRO5 is number of five-membered rings.

The p$K_{m,app}$ was permuted 100 times, and models were generated for each permution. No model resulted in as good a fit ($r^2$) as the original, indicating that the model is highly unlikely ($p < 0.01$) to be a result of a chance correlation. Furthermore, the average "leave *N* out" cross-validated $r^2$ of 100 trials with *N* substrates omitted ($N = 2-7$) was performed to check that the model is likely to generalize to novel chemicals, and the results are summarized in Table 4. The models made with these descriptors had good CV $r^2$ values, between 0.79 ($n = 2$, 11% of the set) and 0.75 ($n = 7$, 39% of the set), and showed low variability indicating a reliable model suited to generalization.[19]

## Discussion

Determination of a pharmacophore represents an early, but significant, step toward the understanding of a given receptor−ligand binding event. However, while satisfaction of the pharmacophore criteria by a molecule is necessary for a binding interaction to occur, it may not be sufficient because many of the steric and electronic influences on the binding event are ignored in such an approach. Therefore, a 3D-QSAR modeling approach that more specifically investigates such influences, such as SOMFA, could complement the pharmacophore approach well. Still, there are other molecular determinants left unexplored by such approaches; hence, a 2D-QSAR approach is potentially beneficial because it allows exploration of an even wider range of molecular determinants that may explain observed phenomena. Thus, the aim of the research reported here was to determine whether a complementary range of models could be developed that describe the features necessary for a compound to undergo glucuronidation by UGT1A4 and aid in the prediction of $K_{m,app}$ for putative substrates. Previous 2D- and 3D-QSAR models of UGT substrates/inhibitors have not been generally applicable because they have used compounds of low structural diversity (usually with a common substructure), making extrapolation to structurally diverse datasets virtually impossible.[20−28]

Substrates of phase I and II enzymes share a binding mode whereby the atom at which reaction occurs must be located at a specific position in the active site for catalysis to occur. For substrates of UGT, this site is adjacent to the bound cofactor UDP-glucuronic acid. Hence, the minimum requirement for the alignment (i.e., pharmacophore) is the mutual overlay of the reactive site for each substrate. Indeed, the routine incorporation of a "site of reactivity" feature into other enzyme-based pharmacophore investigations should generally aid in the alignment process. Such an approach decreases the complexity of the "pharmacophore space" to be searched, thereby increasing the chance of finding a chemically sensible alignment between substrates. Successful pharmacophore developments for CYP have included the site of oxidation as a pharmacophore feature. For example, de Groot et al anchored the substrate oxidation site 3 Å from the iron atom.[29,30] However, the reported CYP pharmacophores developed using Catalyst[7−12] make no mention of whether the alignments from the pharmacophores result in the overlay of a common site of metabolism. This deficiency was addressed recently in the development of a pharmacophore model of CYP2B6, using Catalyst, which incorporated the site of oxidation as a pharmacophore alignment feature.[14] It should be noted, however, that this model cannot be used to search a database of potential substrates because the "site of reactivity" feature is defined only for the training set of compounds and will not recognize the site of reactivity in other chemicals.

There is general acceptance that UGT-catalyzed conjugation proceeds by a second-order nucleophilic substitution ($S_N2$) mechanism[31] and hence requires a nucleophilic site within a molecule. Furthermore, Yin and co-workers propose that a basic residue within the active site may deprotonate aglycons with a p$K_a$ high enough to require it (e.g., a phenol) prior to reaction.[27] Although sufficient kinetic data are not available, the concept of this mechanism can be expanded to substrates with lower p$K_a$ that are sufficiently nucleophilic that they do not require initial deprotonation but do require deprotonation of the intermediate transition state. This hydrogen-bonding interaction potentially represents a significant contribution in binding affinity to the protein, thus justifying the inclusion of a "glucuronidation feature" based on a HBD site in the pharmacophore. Hence, for a site to undergo UGT-catalyzed conjugation, the *general* requirement is a nucleophilic heteroatom attached to a proton (e.g., SH, OH, NH). There are, however, exceptions that must be borne in mind when considering the mechanism by which UGTs function. Examples such as tertiary amine or carbon glucuronidation cause a paradox in relation to describing a general pharmacophore feature that represents the glucuronidation site. HBDs do not accurately represent tertiary amines, and HBAs do not accurately represent C-based glucuronidation sites. In general, however, a HBA feature (with its implicit lone pair) is probably the most representative.

The "common features" pharmacophore allows the inference that the general features responsible for the binding of UGT1A4 substrates with a $K_{m,app}$ between 5 and 500 $\mu$M are two hydrophobic regions and a region

occupied by the glucuronidation site, in the geometry highlighted in Figure 2a. Hence, the pharmacophore provides a useful tool for gaining insight into the UGT active site in the absence of a crystal structure. However, these general features do not necessarily explain all the interactions of higher affinity substrates. For example, it is likely that steroids benefit from a second hydrogen-bonding interaction at the ring most distal to the site of glucuronidation. It was noted that a number of higher affinity substrates shared a relatively common region of HBA groups near the 6.5 Å hydrophobe when aligned, although not within sufficient tolerance to allow Catalyst to assign a feature to the region. It is hypothesized that an HBA feature within 1−2 Å of this hydrophobic site may account for some of the selectivity of the higher affinity substrates. It has been recognized for some time that substrate hydrophobicity (lipophilicity) is essential for glucuronidation by hepatic UGTs.[25−27,31] Interestingly, the interfeature distances highlighted (3−4 Å and 6.2−6.5 Å) in the "common features" pharmacophore parallel those proposed for rat UGT1A6.[32] For the first time, this study sheds light on the geometry of the hydrophobic regions within the substrates and hence provides the beginning of a deeper, mechanistic understanding of the human UGT isoforms. Pharmacophore models developed concurrently for UGT1A1 in this laboratory indicate that this isoform may share a similar set of common features.[33]

Although not quantitative, the "common features" model has potential for qualitative predictions of the preferred site of conjugation in molecules with multiple glucuronidation sites. For example, clozapine is known to have two sites of conjugation: one secondary site and one tertiary amine site.[15] It is interesting to note that this model correctly predicts the major conjugation site (secondary amine) (see Figure 2c) as evidenced by the default "fast fit" alignment on the "glucuronidation" feature of the pharmacophore. Several of the small molecules (for menthol, carvacrol, eugenol, $500 < K_{m,app} < 1600$ $\mu$M) also fit the model very well (e.g., Figure 2b) despite being excluded from the training set used to develop the model. In addition to potentially accounting for multiple binding modes, this adds further credence to the model regarding the minimum features necessary for binding.

Of the predictive models, the 2D-QSAR exhibited the best overall predictivity for the chosen test set. To further investigate the robustness of the 2D-QSAR model for prediction, it is useful to test with more than one test set. One method for doing this is to split the dataset into training and test sets in multiple ways and to determine the average prediction accuracy of the multiple test sets. Leave-*N*-out cross validation allows this and is, in essence, a validation with hundreds of test sets of differing composition. Hence, our 2D-QSAR model was further assessed by determining the average "leave-*N*-out" cross-validated $r^2$ of 100 trials with $N$ substrates omitted ($N = 2-7$). Table 4 highlights these data and the associated standard deviations. The CV $r^2$ values were good, between 0.80 ($n = 2$, 11% of the set) and 0.75 ($n = 7$, 39% of the set), and showed low variability indicating good prediction independent of the particular test set chosen. These values better represent the expected outcome[18] when predicting the activities of other chemicals than does a leave-one-out cross-validated $r^2$. A direct comparison of these results, unlike the single test set results, cannot be made with the 3D-QSAR models because *n*-fold cross-validation is not implemented in the software.

Qualitative interpretation of the descriptors in the 2D-QSAR model is not simple because many molecular descriptors represent a number extracted by a well-defined algorithm from a molecular representation of a complex system.[34] However, log *P* (AlogP98) indicates that substrate hydrophobicity is important for activity. The importance of substrate hydrophobicity may be due to properties of the active site of the enzyme and/or the membrane environment of the enzyme. Since UGTs are believed to be located on the lumenal face of the microsomal membrane, substrates must traverse the endoplasmic reticulum to gain access to the active site.[31] Inclusion of the descriptor nRO5 (the number of five-membered rings) may be due to the high-activity steroids present in the dataset but may contain "hidden" information relating to hydrophobicity also. Physical interpretation of the BCUT descriptor (BENV1) and the two 2D autocorrelation descriptors (MATS3V, MATS5e) is more difficult. It is noteworthy, however, that BCUT descriptors have been shown previously to incorporate connectivity information and atomic properties relevant to receptor−ligand intermolecular interactions.[35] Furthermore, the autocorrelation descriptors encapsulate volume (MATS3V) and electronegativity (MATS5e) information associated with atom pairs separated by three and five bonds, respectively.[34]

Two 3D-QSARs were generated using Catalyst and SOMFA. The same sets of pharmacophoric features (two hydrophobes and a glucuronidation feature) were elucidated using the Catalyst Hypogen module (3D-QSAR) as were elucidated using HipHop, although the geometry was different. Hence, the discrimination was achieved on the basis of the differing "fit" of molecules as opposed to differing features of the highly specific substrates vs the low-specificity substrates. The Hypogen model showed a high correlation ($r = 0.94$; $r^2 = 0.89$) between predicted and experimental $K_{m,app}$ and was able to fit the 18 molecules in the training set to within 0.3 log unit of the experimentally determined $K_{m,app}$. Table 1 shows the predicted $K_{m,app}$ for the test set of six molecules using the Hypogen model, five of which were predicted to be within 1 log unit. This currently represents the level of predictivity being achieved by most literature reports within the drug metabolism field.[10] It should be noted that some of the mappings for the predictive pharmacophore-based 3D-QSAR do not represent catalytically "sensible" mappings (e.g., Figure 3c) because the site of conjugation is not mapped to the glucuronidation function of the pharmacophore. For example, by use of Catalyst's default "Fast Fit" method, imipramine maps to all three features whereas benzidine maps preferentially only to the two hydrophobes, omitting the glucuronidation site. The Hypogen algorithm searches for structural differences to explain differences in activity, thus causing it to ignore the common scaffold when assigning pharmacophoric features.[36] Hence, a situation arises whereby in order to account for some feature-poor or small compounds that bind with moderate affinity, it is necessary for the

Hypogen algorithm to allow compounds to map with high precision but to only two of the three features. For example, using Catalyst's default "Fast Fit" method, imipramine ($K_{m,app} = 310\ \mu M$) maps to all three features but poorly, whereas small, feature-poor benzidine ($K_{m,app} = 460\ \mu M$) maps very precisely but only to the two hydrophobes, omitting the glucuronidation site. This results in a nonsense alignment if judged by the requirement of a pharmacophore to overlay the "site of reactivity" as outlined above. The challenges presented to Catalyst by feature-poor or small compounds with moderate to high affinity are discussed further in a review by Sprague.[37] This is one of the primary reasons for developing a "common features" model that maps features in a "sensible", chemically reasonable manner because it has been built up sequentially and logically. Both models contribute useful, yet differing, information.

Alignment and conformer selection have been highlighted as the most important stage in molecular field-based 3D-QSAR methods;[38] hence, the SOMFA model serves as a partial validation of the "common features" alignment, as well as a model in its own right. The shape-based SOMFA model ($r = 0.85$; $r^2 = 0.73$) implies that the "common features" pharmacophore may provide a useful alignment of the UGT1A4 substrates and additionally may provide useful insights regarding steric interactions that cannot be predicted using only a pharmacophore model. The predictive ability of the SOMFA model, based on the six test chemicals, was comparable with that of the Hypogen model, although both were less accurate in prediction than the 2D QSAR.

As noted above, models generated for CYP have generally achieved prediction similar to that reported here (i.e., 1 log unit). However, increased prediction accuracy is clearly desirable. Important factors with regard to increasing prediction accuracy include (1) a systematic search of existing and new modeling techniques to find those most suitable for UGT and (2) an increase in the availability of isoform substrate/inhibitor selectivity and regioselectivity data. In addressing the first factor, this study provides a comparison of the major types of substrate-based modeling techniques in current use and provides the basis for ongoing studies in the field. While the linear 2D- and 3D-QSAR methods reported here show usefulness, it is possible that nonlinear pattern recognition techniques that are capable of accounting for data from more complex relationships will prove to be more appropriate. Research is currently underway in these laboratories to explore this possibility in light of the understanding gained from this study. It is also anticipated that the accuracy of predictive UGT models will further improve as increasing numbers of chemicals are screened against these enzymes.

## Conclusion

In conclusion, we have reported here a generalized, complementary approach to the challenge of modeling substrates of UGT isoforms. The work provided important insights into the structural and chemical properties that characterize substrates of UGT1A4 and has allowed an evaluation of several pharmacophore, 2D-, and 3D-QSAR pattern recognition techniques for the quantitative prediction of binding affinity ($K_{m,app}$). In particular, the common features pharmacophore provides a chemically intuitive representation of the spatial characteristics of known UGT1A4 substrates and affords a useful tool for gaining insight into the UGT active site in the absence of a crystal structure. The 2D-QSAR highlights the correlation with substrate hydrophobicity among other descriptors. The development of a novel, user-defined "glucuronidation" feature within Catalyst is proposed as a useful aid to the further development of pharmacophore-based UGT models. In terms of quantiatative predictivity, it would appear that 2D-QSAR is more effective than the 3D-QSAR techniques used in this study for prediction of binding affinity ($K_{m,app}$). The outcomes of this study will aid in directing further iterations of model development in the UGT substrate modeling field and the drug metabolizing field in general.

**Acknowledgment.** The authors gratefully acknowledge Daniel Robinson and the Computational Chemistry Research Group (Oxford University, U.K.) for the use of the SOMFA software and to Milano Chemometrics and QSAR Research Group for use of the Dragon software.

## References

(1) Roberts, S. A. High-throughput screening approaches for investigating drug metabolism and pharmacokinetics [Review]. *Xenobiotica* **2001**, *31*, 557−589.
(2) Ekins, S.; Waller, C. L.; Swaan, P. W.; Cruciani, G.; Wrighton, S. A.; et al. Progress in predicting human ADME parameters in silico [Review]. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 251−272.
(3) Miners, J. O.; Mackenzie, P. I. Drug glucuronidation in humans. *Pharmacol. Ther.* **1991**, *51*, 347−369.
(4) Wrighton, S. A.; Stevens, J. C. The human hepatic cytochromes P450 involved in drug metabolism. *Crit. Rev. Toxicol.* **1992**, *22*, 1−21.
(5) Lewis, D. F.; Dickins, M.; Eddershaw, P. J.; Tarbit, M. H.; Goldfarb, P. S. Cytochrome P450 substrate specificities, substrate structural templates and enzyme active site geometries. *Drug Metab. Drug Interact.* **1999**, *15*, 1−49.
(6) Dai, R.; Pincus, M. R.; Friedman, F. K. Molecular modeling of mammalian cytochrome P450s [Review]. *Cell. Mol. Life Sci.* **2000**, *57*, 487−499.
(7) Ekins, S.; De Groot, M. J.; Jones, J. P. Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome P450 active sites [Review]. *Drug Metab. Dispos.* **2001**, *29*, 936−944.
(8) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; et al. Three- and four-dimensional quantitative structure activity relationship analyses of cytochrome P-450 3A4 inhibitors. *J. Pharmacol. Exp. Ther.* **1999**, *290*, 429−438.
(9) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; et al. Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2D6 inhibitors. *Pharmacogenetics* **1999**, *9*, 477−489.
(10) Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J. S.; Ring, B. J.; et al. Three- and four-dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors. *Drug Metab. Dispos.* **2000**, *28*, 994−1002.
(11) Ekins, S.; Bravi, G.; Ring, B. J.; Gillespie, T. A.; Gillespie, J. S.; et al. Three-dimensional quantitative structure activity relationship analyses of substrates for CYP2B6. *J. Pharmacol. Exp. Ther.* **1999**, *288*, 21−29.
(12) Ekins, S.; Bravi, G.; Wikel, J. H.; Wrighton, S. A. Three-dimensional-quantitative structure activity relationship analysis of cytochrome P-450 3A4 substrates. *J. Pharmacol. Exp. Ther.* **1999**, *291*, 424−433.
(13) Afzelius, L.; Zamora, I.; Ridderstrom, M.; Andersson, T. B.; Karlen, A.; et al. Competitive CYP2C9 inhibitors: Enzyme inhibition studies, protein homology modeling, and three-dimensional quantitative structure−activity relationship analysis. *Mol. Pharmacol.* **2001**, *59*, 909−919.
(14) Wang, Q.; Halpert, J. R. Combined three-dimensional quantitative structure−activity relationship analysis of cytochrome P450 2B6 substrates and protein homology modeling. *Drug Metab. Dispos.* **2002**, *30*, 86−95.

(15) Green, M. D.; Tephly, T. R. Glucuronidation of Amines and Hydroxylated Xenobiotics and Endobiotics Catalyzed by Expressed Human Ugt1.4 Protein. *Drug Metab. Dispos.* **1996**, *24*, 356−363.
(16) Green, M. D.; Bishop, W. P.; Tephly, T. R. Expressed Human Ugt1.4 Protein Catalyzes the Formation of Quaternary Ammonium-Linked Glucuronides. *Drug Metab. Dispos.* **1995**, *23*, 299−302.
(17) Sprague, P. W. Automated Chemical Hypothesis Generation and Database Searching with Catalyst [Review]. *Perspect. Drug Discovery Des.* **1995**, *3*, 1−20.
(18) Robinson, D. D.; Winn, P. J.; Lyne, P. D.; Richards, W. G. Self-organizing molecular field analysis: A tool for structure−activity studies. *J. Med. Chem.* **1999**, *42*, 573−583.
(19) Whitley, D. C.; Ford, M. G.; Livingstone, D. J. Unsupervised forward selection: a method for eliminating redundant variables. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1160−1168.
(20) 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.
(21) Said, M.; Battaglia, E.; Elass, A.; Cano, V.; Ziegler, J. C.; et al. Mechanism of inhibition of rat liver bilirubin UDP-glucuronosyltransferase by triphenylalkyl derivatives. *J. Biochem. Mol. Toxicol.* **1998**, *12*, 19−27.
(22) Naydenova, Z. G.; Grancharov, K. C.; Alargov, D. K.; Golovinsky, E. V.; Stanoeva, I. M.; et al. Inhibition of Udp-Glucuronosyltransferase by 5′-*O*-Amino Acid and Oligopeptide Derivatives of Uridine-Structure−Activity Relationships. *Z. Naturforsch., C: J. Biosci.* **1998**, *53*, 173−181.
(23) 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.
(24) Ghauri, F. Y.; Blackledge, C. A.; Glen, R. C.; Sweatman, B. C.; Lindon, J. C.; et al. Quantitative structure−metabolism relationships for substituted benzoic acids in the rat. Computational chemistry, NMR spectroscopy and pattern recognition studies. *Biochem. Pharmacol.* **1992**, *44*, 1935−1946.
(25) Kim, K. H. Quantitative structure−activity relationships of the metabolism of drugs by uridine diphosphate glucuronosyltransferase. *J. Pharm. Sci.* **1991**, *80*, 966−970.
(26) Resetar, A.; Minick, D.; Spector, T. Glucuronidation of 3′-azido-3′-deoxythymidine catalyzed by human liver UDP-glucuronosyltransferase. Significance of nucleoside hydrophobicity and inhibition by xenobiotics. *Biochem. Pharmacol.* **1991**, *42*, 559−568.
(27) Yin, H. Q.; Bennett, G.; Jones, J. P. Mechanistic Studies of Uridine Diphosphate Glucuronosyltransferase. *Chem.-Biol. Interact.* **1994**, *90*, 47−58.
(28) Antonio, L.; Grillasca, J.; Taskinen, J.; Elovaara, E.; Burchell, B.; et al. Characterization of catechol glucuronidation in rat liver. *Drug Metab. Dispos.* **2002**, *30*, 199−207.
(29) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 4062−4070.
(30) de Groot, M. J.; Ackland, M. J.; Horne, V. A.; Alex, A. A.; Jones, B. C. Novel approach to predicting P450-mediated drug metabolism: Development of a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* **1999**, *42*, 1515−1524.
(31) Radominska-Pandya, A.; Czernik, P. J.; Little, J. M.; Battaglia, E.; Mackenzie, P. I. Structural and functional studies of UDP-glucuronosyltransferases [Review]. *Drug Metab. Rev.* **1999**, *31*, 817−899.
(32) Jackson, M. R.; Fournel-Gigleux, S.; Harding, D.; Burchell, B. Examination of the substrate specificity of cloned rat kidney phenol UDP-glucuronyltransferase expressed in COS-7 cells. *Mol. Pharmacol.* **1988**, *34*, 638−642.
(33) Sorich, M. J.; Smith, P. A.; McKinnon, R. A.; Miners, J. O. Pharmacophore and Quantitative Structure Activity Relationship Modeling of UDP-Glucuronosyltransferase 1A1 (UGT1A1) Substrates. *Pharmacogenetics* **2002**, *12* (8), 635−645.
(34) Todeschini, R., Consonni, V., Eds. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
(35) Gao, H. Application of BCUT metrics and genetic algorithm in binary QSAR analysis. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 402−407.
(36) Vandrie, J. H.; Nugent, R. A. Addressing the Challenges Posed by Combinatorial Chemistry−3d Databases, Pharmacophore Recognition and Beyond. *SAR QSAR Environ. Res.* **1998**, *9*, 1 ff.
(37) Sprague, P. W.; Hoffmann, R. CATALYST Pharmacophore Models and Their Utility as Queries for Searching 3D Databases. In *Computer-Assisted Lead Finding and Optimization: Current Tools for Medicinal Chemistry*; Wiley-VCH: Weinheim, Germany, 1997; pp 223−240.
(38) Cramer, R. D., 3rd; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967

JM020397C