Molecular Properties and CYP2D6 Substrates: Central Nervous System Therapeutics Case Study and Pattern Analysis of a Substrate Database^[S]

Laura K. Chico, Heather A. Behanna,¹ Wenhui Hu, Guifa Zhong, Saktimayee Mitra Roy, and D. Martin Watterson

Center for Drug Discovery and Chemical Biology (L.K.C., H.A.B., S.M.R., D.M.W.) and Department of Molecular Pharmacology and Biological Chemistry (D.M.W.), Northwestern University, Chicago, Illinois; and Guangzhou Institute of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China (W.H., G.Z.)

Received April 21, 2009; accepted July 31, 2009

ABSTRACT:

CYP2D6 substrate status is a critical Go/No Go decision criteria in central nervous system (CNS) drug discovery efforts because the polymorphic nature of CYP2D6 can lead to variable patient safety and drug efficacy. In addition, CYP2D6 is disproportionately involved in the metabolism of CNS drugs compared with other drug classes. Therefore, identifying trends in small molecule properties of CNS-penetrant compounds that can help discriminate potential CYP2D6 substrates from nonsubstrates would allow additional prioritization in the synthesis and biological evaluation of new therapeutic candidates. We report here the conversion of the CNS drug minaprine from substrate to nonsubstrate, as well as the conversion of the related CNS drug minozac from nonsubstrate to substrate, through the use of analog synthesis and CYP2D6 enzyme kinetic analyses. No single molecular property strongly cor-

related with substrate status for this 3-amino-4-methyl-6-phenylpyridazine scaffold, although molecular volume and charge appeared to be indirectly related. A parsed database of CYP2D6 substrates across diverse chemical structures was assembled and analyzed for physical property trends correlating with substrate status. We found that a complex interplay of properties influenced CYP2D6 substrate status and that the particular chemical scaffold affects which properties are most prominent. The results also identified an unexpected issue in CNS drug discovery, in that some property trends correlative with CYP2D6 substrates overlap previously reported properties that correlate with CNS penetrance. These results suggest the need for a careful balance in the design and synthesis of new CNS therapeutic candidates to avoid CYP2D6 substrate status while maintaining CNS penetrance.

The avoidance of CYP2D6-mediated drug metabolism represents an early project management criteria in drug development because of its potential for variable patient safety and drug efficacy arising from genetic polymorphisms and its involvement in the metabolism of many existing drugs (Kramer et al., 2007; Leeson and Springthorpe, 2007). An example of an unwanted, variable patient response result-

ing from CYP2D6 polymorphisms is the response to codeine, which requires CYP2D6 conversion to the active drug morphine. Patients with a “slow metabolizer” phenotype caused by altered CYP2D6 expression or function can experience reduced analgesic effects resulting from diminished morphine production. In contrast, mutations causing a “rapid metabolizer” phenotype increase toxicity risks as excessive levels of morphine can be produced (Kirchheiner et al., 2007). As a result of such variance, there is considerable interest in identifying features that make small molecules favorable CYP2D6 substrates to prioritize early discovery efforts toward compounds with a lower likelihood of CYP2D6 involvement.

The prevailing model of favorable CYP2D6 substrates includes the presence of a basic, protonated nitrogen atom 5, 7, or 10 Å from the site of metabolism (de Groot et al., 1997), but the model cannot accommodate substrates lacking the protonated nitrogen or nonsubstrates that possess the protonated nitrogen (Guengerich et al., 2002).

This work was supported in part by the National Institutes of Health National Institute on Aging [Grants AG000260 and AG031311]; and the National Institutes of Health National Institute of Neurological Disorders and Stroke [Grant NS056051].

¹ Current affiliation: Astellas Research Institute of America, Skokie, Illinois.

Article, publication date, and citation information can be found at <http://dmd.aspetjournals.org>.

doi:10.1124/dmd.109.028134.

[S] The online version of this article (available at <http://dmd.aspetjournals.org>) contains supplemental material.

ABBREVIATIONS: CNS, central nervous system; LogP, log₁₀ of the octanol/water partition coefficient; K_M, Michaelis-Menten constant; HPLC, high-performance liquid chromatography; ESI, electrospray ionization; k_{dep}, depletion rate constant (per minute); [S], substrate concentration (micromolar); MW, molecular weight; LogD_{7.4}, log₁₀ of the distribution coefficient at pH 7.4; PSA, polar surface area (Å²); MV, molar volume (cm³); TP, true positive; TN, true negative; FP, false positive; FN, false negative; Eq, equivalent; MW01-1-085HAB, *N*-(2-morpholinoethyl)-6-phenylpyridazin-3-amine; DMSO, dimethyl sulfoxide; MW01-7-121HAB, 4-methyl-6-phenyl-*N*-(2-(tetrahydro-2*H*-pyran-4-yl)ethyl)pyridazin-3-amine; MW01-7-103HAB, 4-methyl-6-phenyl-*N*-(2-(piperidin-1-yl)ethyl)pyridazin-3-amine; t_r, retention time (minutes); MW01-8-071HAB, 4-methyl-6-phenyl-*N*-(2-(piperazin-1-yl)ethyl)pyridazin-3-amine; MW01-8-019HAB, 4-methyl-6-phenyl-3-(4-(pyridin-3-yl)piperazin-1-yl)pyridazine; MW01-2-068ZGF, *N*¹,*N*²-dimethyl-*N*¹-(4-methyl-6-phenylpyridazin-3-yl)-*N*²-(pyrimidin-2-yl)ethane-1,2-diamine.

Characterization of what makes a particular small molecule a CYP2D6 substrate is often done through examination of how structural changes alter discrete properties and CYP2D6 substrate status (Uphagrove and Nelson, 2001; Kalgutkar et al., 2003). Such investigations provide insight into the features influencing CYP2D6 substrate status for specific sets of small molecules, but the trends cannot necessarily be generalized across other small molecule drug candidates. In the absence of general models that can be applied with confidence to any small molecule drug discovery campaign, ongoing projects must synthesize and experimentally screen all possible candidates for potential CYP2D6 substrate status.

The importance of first-pass metabolism of drugs by CYP2D6 is of greater impact for central nervous system (CNS) drugs (Goodwin and Clark, 2005), where inspection of the DrugBank database (Wishart et al., 2008) reveals that 31% of approved CNS drugs are metabolized by CYP2D6. For this reason, it is essential to evaluate potential trends among the available approved CNS-penetrant drugs and experimental therapeutics. However, CNS drugs constitute only a small percentage ($\leq 5\%$) of approved therapeutics (Wishart et al., 2008), making the detection of statistically significant trends difficult. As an alternative, analysis of structurally related small molecule CNS drugs has the potential to test prevailing hypotheses and provide insight. In this regard, minaprine and minozac provide a novel study of structurally similar compounds (Tanimoto similarity coefficient = 0.71) with distinct pharmacology and CNS activity that are also different in their CYP2D6 substrate status. In terms of prevailing models, the difference in CYP2D6 status of minaprine and minozac cannot be explained by differences in lipophilicity (de Groot et al., 1997) because the compounds possess nearly identical experimental $\log_{10}$ of the octanol/water partition coefficient (LogP) values of 1.9 (Arnaud-Neu et al., 1990) and 2.3 (Hu et al., 2007), respectively. Although minaprine adheres to the prevailing localized charge model for CYP2D6 substrate status (de Groot et al., 1997), in that it is more basic than the nonsubstrate minozac, and a quantitative model of CYP2D6 substrates provides an accurate forecast for minaprine (Haji-Momenian et al., 2003), the quantitative structure-activity relationship model cannot explain why minozac is not a substrate.

To gain insight into the molecular basis of the CYP2D6 substrate differences for minaprine and minozac, we sought to interconvert minaprine from a substrate to a nonsubstrate and minozac from a nonsubstrate to a substrate. This experiment was done by synthesis of analogs for each CNS drug and evaluating the CYP2D6 enzyme kinetics for each compound. The trend in molecular properties when a substrate is transformed into a nonsubstrate (and vice versa) does not follow a simple correlation with any given single molecular property but suggests an indirect effect of charge and molecular volume. As part of an effort to place these results into a larger context, we generated a parsed database of CYP2D6 substrates from the available literature and subjected it to database mining analysis. The analyses identified complex relationships between CYP2D6 Michaelis-Menten constant (K_M) values and molecular properties related to charge, lipophilicity, and molecule size, with the pattern of properties depending on the particular chemical scaffold being examined, consistent with the outcomes from the comparative evaluation of minaprine and minozac.

Materials and Methods

Chemicals and Reagents. High-performance liquid chromatography (HPLC)-grade formic acid, minaprine, dibasic potassium phosphate, and quinidine were purchased from Sigma-Aldrich (St. Louis, MO). Perchloric acid was purchased from Thermo Fisher Scientific (Waltham, MA). Chemicals and solvents used for syntheses were purchased from Aldrich Chemical Co. (Mil-

waukee, WI) and used as received unless otherwise stated. Monobasic potassium phosphate was purchased from Mallinckrodt (Hazelwood, MO). HPLC-grade acetonitrile was purchased from Honeywell Burdick & Jackson (Morristown, NJ). Water was from a Millipore Corporation (Billerica, MA) Milli-Q Biocel A10 system.

Analytical Chemistry. All the HPLC analytical data were obtained on a commercially available system from Dionex Corp. (Sunnyvale, CA) by using a Phenomenex (Torrance, CA) Luna C18 column (250×2.0 mm; $5 \mu\text{m}$) with guard column as described previously (Hu et al., 2007). Peak quantification was done based on absorption measurements at 260 nm relative to a standard curve obtained by serial dilutions of compounds. The mobile phase consisted of 0.1% (v/v) formic acid in water as reagent A and acetonitrile with 0.1% (v/v) formic acid in water as reagent B at a flow rate of 0.2 mL/min.

Final compounds were characterized by electrospray mass spectroscopy (ESI), HPLC, and ^1H NMR. All the intermediates were characterized minimally by ESI and HPLC. NMR spectra were acquired on a Varian, Inc. (Palo Alto, CA) Inova 500-MHz spectrometer at room temperature. ESI mass spectra were collected on a Micromass Quattro II Triple Quadrupole HPLC/tandem mass spectrometer.

Determination of Michaelis Constants (K_M) by Substrate Depletion Approach. Kinetics experiments were performed as described previously (Obach and Reed-Hagen, 2002). In brief, compounds ($1\text{--}100 \mu\text{M}$ final concentration) were incubated in triplicate with CYP2D6 (0.02 mg/mL) in a total volume of 1.4 mL. At times 0, 2, 5, 10, 20, and 30 min after initiation of the reactions by the addition of NADPH, 0.2-mL aliquots were removed and added to 70% (v/v) perchloric acid to terminate the reaction. The mixtures were centrifuged at 12,000g for 5 min, and the supernatant was analyzed by HPLC. The percentage of compound remaining versus time for each substrate concentration was fitted to a first-order decay function to determine the initial substrate depletion rate constant (k_{dep}). K_M was determined by plotting the k_{dep} versus substrate concentration on a linear-log plot using the following equation: $k_{\text{dep}} = k_{\text{dep}([S]=0)} \times [1 - ([S]/([S] + K_M))]$, where $[S]$ is defined as the substrate concentration and $k_{\text{dep}([S]=0)}$ is the theoretical maximum consumption rate constant at an infinitesimally low-substrate concentration.

Because of technical limitations with the substrate depletion approach, apparent K_M values are reported here between 0.2 and $30 \mu\text{M}$ based on the experimental conditions. For example, substrate concentrations less than $0.075 \mu\text{M}$ were not evaluated because of limits of reproducible HPLC detection. The determination of K_M values beyond $30 \mu\text{M}$ requires the analysis of higher concentrations of substrates ($\geq 100 \mu\text{M}$), but aqueous solubility limitations prevented the analysis of substrate concentrations greater than $100 \mu\text{M}$. Thus, values are reported as $\geq 30 \mu\text{M}$ where appropriate.

Determination of pK_a Values. The pK_a values for analogs were determined by pION, Inc. (Woburn, MA) using a validated potentiometric method (Avdeef et al., 1993). In brief, stock solutions of test compounds were prepared at a concentration of 20 mg/mL in dimethylsulfoxide. The pK_a was determined in a mixture of water and cosolvent with a minimum of three ratios of water/cosolvent titrated to obtain an apparent pK_a in the presence of cosolvent (p_sK_a). The potentiometric pK_a was determined by extrapolation of the p_sK_a values measured in the cosolvent titrations using the Yasuda-Shedlovsky technique (Avdeef et al., 1993).

CYP2D6 Kinetics Database. A database of small molecules with reported experimental CYP2D6 kinetics data was generated from literature sources using SciFinder Scholar (Supplemental Data Excel File). Keyword searching used the terms "CYP2D6," "substrate," and "kinetics" to identify publications that were manually inspected for relevant data. Inclusion in the database was limited to compounds that had reported K_M obtained using purified CYP2D6 enzyme systems. The following molecular properties for each database member were calculated using ACD/Labs version 11.0 (Advanced Chemistry Development, Toronto, ON, Canada): molecular weight (MW), lipophilicity (LogP and LogD_{7.4}), polar surface area (PSA), hydrogen bond donor count, hydrogen bond acceptor count, freely rotatable bonds, highest computed pK_a value (high- pK_a), polarizability, molar volume (MV), and number of rings. These properties were selected because of previous reports of involvement in mediating CYP2D6 substrate recognition. Chemical Abstracts Registry numbers (CAS number), chemical structure (SMILES), and K_M values were also noted for each compound.

Descriptive statistical analyses of molecular property trends among

CYP2D6 substrates were performed using the statistical software package Prism version 4.0 (GraphPad Software Inc., La Jolla, CA). Statistical correlations between normalized kinetic values ($\log_{10}[1/K_M]$) and individual computed molecular properties were determined for the CYP2D6 substrate database using Prism version 4.0 (GraphPad Software Inc.). Differences in mean property values between substrates and nonsubstrates were evaluated for statistical significance using the Student's *t* test in Prism version 4.0, where $* = p < 0.05$ and $** = p < 0.01$.

Construction of Decision Tree Models. The data-mining software WEKA version 3.4.12 (Witten and Frank, 2005) was used to construct J48 decision trees that forecast CYP2D6 substrate status from an automatically selected subset of computed molecular properties. The J48 decision tree is an implementation of the C4.5 algorithm (Quinlan, 1993). Compounds were classified as substrates if the K_M value was less than or equal to 25 μM , whereas nonsubstrates were considered to be compounds with K_M values greater than 25 μM . The selection of this cutoff was based on the kinetics values of favorable CYP2D6 substrates identified from literature sources and the experimental studies with minaprine and minozac. For example, bufuralol (CAS number 54340-62-4) and dextromethorphan (CAS number 125-71-3) possess K_M values equal to or less than 15 μM (Yu, 2001; Haji-Momenian et al., 2003). These compounds are routinely used as standard CYP2D6 substrates in enzyme assays and are also recommended by the Food and Drug Administration to be used in the evaluation of novel preclinical and clinical candidates. For the purposes of this study, compounds with K_M values less than 25 μM were categorized as substrates to ensure that a range of compounds serving as moderate to favorable substrates were classified separately from nonsubstrates.

Molecular properties for compounds were computed as described above using ACD/Labs software and were manually transformed from a tab-delimited file format to the required input ARFF file format (format described at <http://www.cs.waikato.ac.nz/~ml/weka/arff.html>). The J48/C4.5 decision tree algorithm performs an automated variable selection to identify decision trees with the most statistically meaningful descriptors. However, it is important to note that statistical significance does not necessarily imply scientific significance; thus, thorough inspection and interpretation of the resulting decision trees were necessary.

Models were initially constructed using a 10-fold cross-validation technique to assess potential overfitting. In *n*-fold cross-validation, the supplied data set is randomly partitioned into *n* groups, and one partition is withheld from construction and used to test the model, and the remaining partitions are used for model construction. Model rebuilding is performed *n* times with a new partition withheld each cycle to allow all the compounds in the input set to be used at least once in testing.

To assess the potential impact of different chemical scaffolds on resulting model structure, we created three different models described in the Supplemental Data Information File. Each model was built on a differently randomized subset of compounds in the CYP2D6 substrate database. To identify a construction or training set of compounds and a validation or testing set, all the database members were randomly assigned a number from 0 to 1. Compounds with a random number greater than or equal to 0.7 were assigned to the test set, whereas the other compounds were retained in the training set. Compounds in the test set were not used in model construction. This process was repeated three different times to generate three different training and test sets.

Model performance was primarily evaluated using classification accuracy, sensitivity, and specificity as defined below in equations 1 through 3. True positive (TP) and true negative (TN) refer to compounds that are correctly assigned as a substrate or nonsubstrate, respectively. Likewise, false positive (FP) and false negative (FN) refer to compounds that are incorrectly assigned as substrate or nonsubstrate, respectively.

1. Accuracy = $(TP + TN)/(TP + FN + TN + FP)$
2. Sensitivity = $TP/(TP + FN)$
3. Specificity = $TN/(TN + FP)$

Synthesis. Minozac, MW01-5-079HAB, and MW01-5-042HAB were synthesized and characterized as described previously (Hu et al., 2007). Synthetic reaction progress was monitored by HPLC, and final products were characterized by mass spectrometry, HPLC, and ^1H NMR as described previously (Hu et al., 2007). Yields of final compounds were determined gravimetrically.

General Procedure for the Synthesis of Aminated Pyridazines. Two general methods (A and B) were used for the synthesis of aminated pyridazines. Method A: a mixture of the appropriate chloropyridazine precursor (1 Eq) and amine in solvent was reacted by microwave irradiation (CEM Discover, Matthews, NC) in a capped 10-ml microwave glass vessel. Microwave irradiation of 125 W was used, ramping from ambient temperature to 175°C. On reaching the set temperature of 175°C, the reaction mixture was held for 3 h, allowed to cool to ambient temperature and purified to obtain product. Method B: the appropriate chloropyridazine precursor (1 Eq) and amine were refluxed in 1-butanol under argon gas. The solvent was removed under reduced pressure and further purified to afford the final product.

MW01-1-085HAB. 2-Morpholinoethanamine (1.3 g, 10.5 mmol) and 3-chloro-6-phenyl-pyridazine (500 mg, 2.62 mmol) were reacted according to Method B. Following evaporation, the brown residue was washed with water and 0.1% acetic acid. The precipitate was filtered, dried, and collected in 35% yield as a pale white, crystalline powder. ^1H NMR [dimethyl sulfoxide (DMSO)]: δ 7.98 (m, 2H); 7.60 (d, $J = 9$ Hz, 1H); 7.47 to 7.39 (m, 3H); 6.57 (d, $J = 9$ Hz, 1H); 5.35 (bs, 1H); 3.74 (t, $J = 4.5$ Hz, 4H); 3.60 (m, 2H); 2.70 (t, $J = 6$ Hz, 2H); 2.53 (bs, 4H). HPLC (t_r , purity): 15.77, 99%; ESI m/z (MeOH): 285.20 (MH^+).

MW01-7-121HAB. 2-(Tetrahydro-pyran-4-yl)ethylamine (813 mg, 5.86 mmol) and 3-chloro-4-methyl-6-phenylpyridazine (300 mg, 1.47 mmol) were reacted in water according to Method A. The product was purified by solid-phase extraction using equilibrated Waters (Milford, MA) Sep-Pak C18 cartridges and eluted with 5% (v/v) acetonitrile in 0.1% formic acid (v/v) to produce a pale orange oil in 35% yield. ^1H NMR ($\text{DMSO}-d_6$): δ 7.96 (d, $J = 7.5$ Hz, 2H); 7.68 (s, 1H); 7.44 (t, $J = 7.5$ Hz, 2H); 7.36 (t, $J = 7.5$ Hz, 1H); 6.33 (t, $J = 5.5$ Hz, 1H); 3.82 (dd, $J_1 = 3.5$ Hz, $J_2 = 11$ Hz, 2H); 3.52 (m, 2H); 3.27 (t, $J = 11.2$ Hz, 2H); 2.13 (s, 3H); 1.64 to 1.57 (m, 5H); 1.19 (bs, 2H). HPLC (t_r , purity): 15.08, 99%; ESI m/z (MeOH): 298.29 (MH^+).

MW01-7-103HAB. *N*-(2-Aminoethyl)piperidine (1.32 ml, 9.19 mmol) and 3-chloro-4-methyl-6-phenylpyridazine (470 mg, 2.30 mmol) were reacted in water according to Method A. On cooling, the water was decanted from the mixture, and the remaining oil was subsequently dissolved in methylene chloride, extracted with water, and dried over sodium sulfate. The solvent was evaporated under reduced pressure to produce a yellow oil that was triturated with ether to give a yellow solid collected and dried in vacuo to afford product in 15% yield. ^1H NMR ($\text{DMSO}-d_6$): δ 7.97 (d, $J = 8.0$ Hz, 2H); 7.70 (s, 1H); 7.45 (t, $J = 7.5$ Hz, 2H); 7.37 (t, $J = 7.0$ Hz, 1H); 6.23 (bs, 1H); 3.59 (m, 2H); 2.53 (t, $J = 7$ Hz, 2H); 2.40 (bs, 4H); 2.13 (s, 3H); 1.50 (m, 4H); 1.38 (bs, 2H). HPLC (t_r , purity): 11.83 min, 95%; ESI m/z (MeOH): 297.22 (MH^+).

MW01-8-071HAB. 4-*N*-(2-aminoethyl)-*N*-Boc-piperazine (896 mg, 3.91 mmol) and 3-chloro-4-methyl-6-phenylpyridazine (200 mg, 0.98 mmol) were reacted in ethanol according to Method A. The temperature was held at 110°C to avoid thermal cleavage of the Boc protecting group, and the irradiation cycle was repeated five times to reach 50% conversion to product. The reaction was cooled to ambient temperature, and the ethanol was evaporated under reduced pressure. The product was separated from residual pyridazine-starting material by column chromatography, with the product eluting with 5% (v/v) methanol in methylene chloride. The compound was collected as a yellow oil (50% conversion, 80% yield), confirmed by mass spectrometry [ESI, 398.25 (MH^+)] and used without further purification. The oil was treated with 3 M HCl in ethyl acetate for 30 min. The reaction mixture was evaporated under reduced pressure, washed with ether, and collected as a yellow powder in 85% yield. ^1H NMR (CDCl_3): δ 8.28 (s, 1H); 7.94 (d, $J = 7$ Hz, 2H); 7.67 (m, 3H); 4.05 (t, $J = 5$ Hz, 2H); 3.76 (bs, 4H); 3.70 (t, $J = 5$ Hz, 4H); 3.31 (s, 2H); 2.60 (s, 3H); 2.02 (s, 1H). HPLC (t_r , purity): 8.93 min, 97%; ESI m/z (acetonitrile): 298.31 (MH^+).

MW01-8-019HAB. 1-Pyridin-3-yl-piperazine dihydrochloride (703 mg, 2.93 mmol) and 3-chloro-4-methyl-6-phenylpyridazine (200 mg, 0.98 mmol) were reacted in water according to Method A. Triethylamine (8 Eq, 1.1 ml) was added to the reaction mixture to neutralize the piperazine salt. The reaction was cooled to ambient temperature, diluted with water, and extracted, dried, and solvent evaporated. The crude product was purified by column chromatography and eluted as a yellow oil with 3% (v/v) methanol in methylene chloride. The oil was then dissolved in methylene chloride/hexanes (1:1) and placed at 4°C to allow formation of product as yellow crystals in 19% yield. ^1H NMR (CDCl_3): δ 8.37 (s, 1H); 8.14 (s, 1H); 8.04 (d, $J = 7.0$ Hz, 2H); 7.60

TABLE 1
CYP2D6 K_M and pK_a values of minaprine and minozac and their analogs

<div> </div>									
Compound	R ₃	R ₄	K _M	pK _a Values		LogP	LogD _{7.4}	PSA	MV
				Experimental	Computed				
Minaprine		CH ₃	1.45 ± 0.43 <i>μM</i>	4.27 ± 0.12 , 7.23 ± 0.12	4.63 ± 0.50 , 7.09 ± 0.10	1.45	1.83	50.28	257.99
MW01-1-085HAB		H	5.01 ± 0.18	3.74 ± 0.03^a , 6.74 ± 0.01^a	3.86 ± 0.50 , 7.09 ± 0.10	2.39	2.19	50.28	241.71
MW01-8-071HAB		CH ₃	>30	3.90 ± 0.24 , 8.65 ± 0.24	3.33 ± 0.50 , 5.06 ± 0.50 , 9.16 ± 0.10	1.87	−0.02	53.08	262.98
MW01-7-103HAB		CH ₃	0.21 ± 0.28	4.42 ± 0.03 , 9.41 ± 0.12	4.68 ± 0.50 , 8.72 ± 0.10	3.62	2.25	41.05	266.94
MW01-7-121HAB		CH ₃	2.70 ± 1.90	5.61 ± 0.18	5.15 ± 0.50	2.80	2.79	47.04	266.92
Minozac		CH ₃	>30	3.75 ± 0.06	4.87 ± 0.33 , 7.33 ± 0.50	1.98	1.52	58.04	271.40
MW01-5-042HAB		CH ₃	0.72 ± 0.36	7.83 ± 0.12	6.37 ± 0.42 , 7.33 ± 0.50	1.45	0.91	32.26	240.76
MW01-8-019HAB		CH ₃	>30	3.79 ± 0.06 , 5.79 ± 0.09	6.97 ± 0.12 , 7.33 ± 0.50	2.07	1.56	45.15	278.18
MW01-2-068ZGF		CH ₃	>30	2.37 ± 0.09 , 4.56 ± 0.06	3.24 ± 0.50 , 5.76 ± 0.31	4.40	4.39	58.04	280.77
MW01-5-079HAB		CH ₃	>30	8.58 ± 0.12	6.68 ± 0.50 , 7.66 ± 0.70	3.42	2.91	32.26	300.88

^a Previously reported value (Arnaud-Neu et al., 1990).

(s, 1H); 7.49 (t, $J = 7.0$ Hz, 2H), 7.45 (d, $J = 7.0$ Hz, 1H); 7.27 (m, 1H); 7.21 (m, 1H); 3.55 (s, 4H); 3.42 (s, 4H); 2.42 (s, 3H). HPLC (t_R /purity): 15.48 min, 99%; ESI m/z (acetonitrile): 332.24 (MH^+).

MW01-2-068ZGF. 3-Chloro-4-methyl-6-phenylpyridazine (1.0 g, 4.9 mmol) and N_1,N_2 -dimethylethane-1,2-diamine (1.3 g, 12.2 mmol) were reacted according to Method B for 72 h. After evaporation, the brown-red residue was treated with water (30 ml) and dichloromethane (30 ml). After extraction with dichloromethane (3×30 ml), the organic layer was dried with sodium sulfate, solvent removed under reduced pressure, and the residue purified by silica gel chromatography (methanol/dichloromethane, 1:20) to give N^1,N^2 -dimethyl- N^1 -(4-methyl-6-phenylpyridazin-3-yl)ethane-1,2-diamine in 72% yield.

The resulting amine (0.90 g, 3.51 mmol) was placed in a 38-ml pressure vessel followed by addition of 2-bromopyrimidine (0.52 g, 3.51 mmol), TEA (1.5 ml, 10.5 mmol), and 1-butanol (8 ml). The vessel was sealed tightly, placed in an oil bath, and stirred at 120°C for 24 h. After cooling to ambient temperature, the reaction mixture was transferred to a single-neck flask for

evaporation under reduced pressure. Removal of solvent gave rise to a brown-red residue that was treated with water (20 ml) and dichloromethane (20 ml). After extraction with dichloromethane (3×30 ml), the organic layer was dried with sodium sulfate; the solvent was removed under reduced pressure; and the residue was purified by silica gel chromatography (PE/ethyl acetate, 3:1) to afford compound 8 in 56% yield. ¹H NMR ($CDCl_3$): δ 8.26 (d, $J = 5$ Hz, 2H); 8.02 (m, 2H); 7.48 to 7.40 (m, 4H); 6.44 (t, $J = 4.5$ Hz, 1H); 3.92 (t, $J = 6.5$ Hz, 2H); 3.64 (t, $J = 7$ Hz, 2H); 3.14 (s, 3H); 3.12 (s, 3H); 2.27 (s, 3H). HPLC (t_R /purity): 24.93, 97%; ESI m/z (MeOH): 335.32 (MH^+).

Results

Molecular Basis for Differential CYP2D6 Substrate Status of Minaprine, Minozac. To evaluate the molecular basis and molecular properties associated with minaprine serving as a substrate and the contrasting CYP2D6 stability of minozac, analogs of the two com-

pounds were synthesized and tested for CYP2D6 substrate status. The overall goal was to reverse the respective CYP2D6 substrate status of minaprine and minozac. Because of the similarities in lipophilicity, additional molecular properties reflecting physical characteristics such as charge and size were computed to identify how property differences between the two compounds might relate to CYP2D6 substrate status. Computed properties are often suitable substitutes to experimental properties if there is congruence between computed and experimental values, such as lipophilicity (LogP) in the case of minaprine and minozac. However, a departure between computed and experimental pK_a values warranted experimental determination of pK_a values for each new compound synthesized in this study. These results are summarized in Table 1.

The minaprine R_6 phenyl ring is the major site of metabolism in humans (Davi et al., 1981), which is common to the 3-amino-4-methyl-6-phenylpyridazine scaffold of both drugs. A secondary site of metabolism is the R_4 methyl group, which is also common to the two drugs. However, substitution of the minaprine R_4 methyl group with a hydrogen atom yielded an analog (MW01-1-085HAB; Table 1) with a similar CYP2D6 K_M value and computed physical properties. A third point of minaprine metabolism that represents that only structural difference between the two drugs is the R_3 position amine. Therefore, the R_3 position amines represented the most logical focus to reverse CYP2D6 substrate status within the minaprine and minozac families (Table 1).

The substrate status of the CYP2D6 substrate minaprine was reversed by the substitution of the minaprine R_3 morpholine ring with a piperazine ring to yield an analog (MW01-8-071HAB) stable to CYP2D6. The reversal of CYP2D6 substrate status correlates with a decrease in $\text{LogD}_{7.4}$ (1.83 to -0.02) and an increase in basicity as measured by pK_a (Table 1). Other minaprine analogs (MW01-7-103HAB, MW01-7-121HAB) with modified morpholine rings span a range of experimental pK_a values (nearly 5 units) and computed lipophilicity values (3 log units) but remain CYP2D6 substrates with K_M values less than $5 \mu\text{M}$ (Table 1). Although it was rather straightforward to transform minaprine from a CYP2D6 substrate to a non-substrate with the synthesis and evaluation of a discrete set of analogs, the evaluation of molecular properties failed to reveal a simple correlation between CYP2D6 substrate status and molecular property values.

In contrast to the approach with converting minaprine from a substrate to a nonsubstrate, the goal of the minozac analysis was to convert a nonsubstrate into a CYP2D6 substrate. The same approach of analog synthesis and evaluation of CYP2D6 substrate status was used. Conversion from nonsubstrate to a substrate with a K_M value comparable with minaprine was accomplished by the synthesis of the minozac analog MW01-5-042HAB, in which the pyrimidine ring of minozac was substituted with a smaller methyl functionality (Table 1). MW01-5-042HAB and minozac have similar lipophilicities but are different in terms of their basicity (pK_a), PSA, and MV. However, other minozac analogs with similar variance of pK_a , LogP, or PSA do not exhibit altered CYP2D6 substrate status. Like the results with minaprine, the synthesis and evaluation of a discrete, logical set of minozac analogs allowed the transformation of a CYP2D6 nonsubstrate to a substrate, but this transformation did not correlate with a simple set of computed molecular properties.

Although the interconversions of the 3-amino-4-methyl-6-phenylpyridazine scaffold between substrate to nonsubstrate status failed to correlate with a single molecular property or simple set of molecular properties, the results raise the possibility of a complex or indirect effect of physical properties that are influenced by MV and charge. To further pursue this emerging hypothesis and gain insight

TABLE 2
Descriptive statistics of CYP2D6 substrate database

	LogP	LogD _{7.4}	PSA	MV
Minimum	-1.67	-1.76	3.24	114.10
Maximum	8.32	6.38	190.20	556.00
Mean	3.02	1.54	37.21	255.40
Median	2.94	1.54	35.53	258.70
25th Percentile	1.85	0.34	19.11	210.80
75th Percentile	3.97	2.36	50.28	289.30

FIG. 1. Distribution of computed LogP (A), $\text{LogD}_{7.4}$ (B), PSA (C), and MV (D) values of CYP2D6 kinetics small molecule database.

into its generality for other chemical scaffolds, a more extensive trend analysis of individual molecular properties for diverse chemical scaffolds and CYP2D6 substrate status is needed.

Generation of a CYP2D6 Substrate Database and Molecular Property Pattern Analysis. We generated a small molecule CYP2D6 substrate database from literature sources with estimated K_M values obtained using purified enzyme sources, while excluding data from hepatocytes and pooled microsomes. Because of the database size and the desire to provide a downloadable spreadsheet for use by other investigators, the database is provided in a spreadsheet format as Supplemental Data Information File 2.

Descriptive statistical analyses for each of several molecular properties in the database were performed to examine the mean and median values and detect potential bias in the available data (Table 2). Key features are highlighted in the following sections and in Figs. 1 through 3. First, the database includes CYP2D6 substrates with a range of K_M values from $0.2 \mu\text{M}$ to 14 mM , but the majority of compounds possess CYP2D6 K_M values between 3 and $75 \mu\text{M}$. Second, the LogP and $\text{LogD}_{7.4}$ values for the database are normally distributed (Fig. 1). Third, PSA and MV values are also normally distributed, but a large proportion of the database falls close to the mean values of $36.70 \text{ \AA}^2$ and 250.60 cm^3 , respectively (Fig. 1, C and D).

The database was analyzed for potential correlations between computed molecular properties and the normalized K_M values

FIG. 2. Relationship between experimental CYP2D6 small molecule kinetics values and computed LogP (A), LogD_{7.4} (B), PSA (C), and MV (D) values.

($\log_{10}[1/K_M]$). First, only LogP, LogD_{7.4}, PSA, and MV showed statistically significant correlations with normalized CYP2D6 K_M values of the 12 computed properties in the database. Scatter plots show clusters of property values associated with K_M values less than 25 μM (Fig. 2, dashed lines). Second, CYP2D6 substrates ($K_M < 30 \mu\text{M}$) possess higher lipophilicity, as indicated by LogP and LogD_{7.4}, than nonsubstrates (Fig. 3, A and B). Substrates possess mean LogP values of 3.41, and nonsubstrates possess mean values of 2.36. Third, an interesting trend of PSA toward lower values was associated with CYP2D6 ($K_M < 30 \mu\text{M}$) substrate status (Figs. 2C and 3C). For example, CYP2D6 substrates ($K_M < 30 \mu\text{M}$) possess a mean PSA value of 32.09 $\text{\AA}^2$, whereas nonsubstrates ($K_M > 30 \mu\text{M}$) have a higher mean PSA value of 45.45 $\text{\AA}^2$ (Fig. 3C). Fourth, volume is a relevant molecular property for CYP2D6 given that compounds with similar MWs can occupy divergent three-dimensional spaces, and a cluster of substrates appears (Fig. 2D) with MV values around 150 to 300 cm^3 . CYP2D6 substrates ($K_M < 30 \mu\text{M}$) possess a higher mean MV (265.93 cm^3), whereas nonsubstrates have a lower mean MV (222.19 cm^3).

Data Mining of CYP2D6 Substrate Database for Higher-Order Molecular Property Patterns. Decision trees are a commonly used data-mining strategy that can identify complex patterns in multidimensional data sets to generate classification models (Kingsford and Salzberg, 2008) and can yield models amenable to correlation with experimental endpoints (Kingsford and Salzberg, 2008). Inherent to the approach is the generation of multiple decision trees using the same data set to reveal stable patterns. Therefore, multiple decision trees were built using the CYP2D6 substrate database described above and the WEKA data mining software suite (Quinlan, 1993; Witten and Frank, 2005) to evaluate patterns of molecular properties that might be

associated with CYP2D6 substrates with this stage of the database. As detailed in the Supplemental Data, the decision trees were constructed from an automatically selected subset of computed molecular properties that best correlated with CYP2D6 substrate status, while minimizing intercorrelation with each other. One limitation of decision tree models can be the sensitivity to the composition of the training set. For example, if the data present in the training set were dominated by compounds with high LogP values, the resulting model will probably contain this as a major classification feature. Therefore, we used differently randomized sets of compounds from the database to build and computationally validate the decision trees to examine potential effects of different chemical scaffolds on model outcomes.

The resulting models varied slightly in terms of their overall appearance (Supplemental Data Information File 2). Two computed properties of CYP2D6 substrates consistently appeared in all the models. Both the number of rings and lipophilicities were features consistent with CYP2D6 substrates in the three decision tree models shown. It is interesting to note that minaprine and minozac are different in their respective ring counts but possess similar lipophilicities. Furthermore, the ring differences between minaprine and minozac contribute to charge differences. Minaprine is a more basic compound with a pK_a value nearly 3.5 units higher than minozac (Table 1). The data-mining results show that simple correlations between singular molecular properties and CYP2D6 substrate status do not necessarily apply across diverse chemical scaffolds, but rather more complex combinations of molecular properties influence CYP2D6 substrate status. In this regard, the 3-amino-4-methyl-6-phenylpyridazine scaffold case study of minaprine and minozac is compliant with the decision tree outcomes for all the known CYP2D6

FIG. 3. Differences in mean molecular property values for CYP2D6 substrates and nonsubstrates. Mean values for LogP (A), LogD_{7.4} (B), PSA (C), and MV (D) were computed for CYP2D6 database members with K_M values less than 30 μ M (substrates) and greater than or equal to 30 μ M (nonsubstrates). Statistical significance of mean differences evaluated using Student's *t* test. *, $p < 0.05$; **, $p < 0.01$.

substrates as it represents a subset of the larger population of chemical scaffolds.

Discussion

Several observations emerge from this study. First, it was possible to interconvert the CYP2D6 substrate status of both minaprine and minozac by making discrete modifications to the R_3 position amines. Second, the case study investigation of minaprine and minozac analogs showed that in cases where simple combinations of molecular properties cannot accurately explain or forecast CYP2D6 substrate status, data mining identifies combinations of properties that collectively influence CYP2D6 substrate status. Third, the results of the CYP2D6 database pattern analysis reported here raise an apparent paradox for CNS drug discovery concerning the influence of molecular properties on CYP2D6 substrate status and small molecule blood-brain barrier penetration. Fourth, the release of the parsed CYP2D6 substrate database reported here provides a resource correlating CYP2D6 K_M values, molecular properties, and small molecule structure that can be downloaded, updated, and used for future refinement and development of CYP2D6 computational tools.

The interconversion of the 3-amino-4-methyl-6-phenylpyridazine scaffold between substrate and nonsubstrate is caused by a complex interplay of physical features that determine CYP2D6 substrate status. For example, volume and basicity changes correlated with CYP2D6 substrate status for several analogs but not for the entire set of compounds. The results show the limited utility of applying simple sets of molecular property descriptors as a CYP2D6-forecasting tool. The comparison of molecular property trends in the minaprine and minozac analog series with patterns emerging from the database of all the CYP2D6 substrates revealed similarities in attributes with a larger

and more diverse collection of chemical scaffolds, but it also indicated that the precise combination of properties affecting CYP2D6 substrate status depends on the chemical scaffold being diversified.

An unexpected issue was raised by this study for researchers interested in minimizing CYP2D6 involvement and developing CNS-penetrant therapeutics. In particular, molecular property guidelines that include minimizing PSA and increasing LogP are often used to improve small molecule brain partitioning (Clark, 1999; Gleeson, 2008). However, the pattern analysis of the CYP2D6 database showed that lower PSA values coincide with an increased likelihood of serving as a CYP2D6 substrate. Likewise, we also identified a pattern of higher LogP values associated with favorable CYP2D6 substrate status. Thus, if the PSA and LogP values of potential drug discovery candidates are optimized to improve brain uptake, there exists a potential undesired consequence of generating favorable CYP2D6 substrates. In this regard, it is interesting to note that the analysis of the DrugBank database reveals that approved CNS drugs are disproportionately CYP2D6 substrates compared with other drug classes. The CYP2D6 substrate database analysis reported here raises the possibility that a critical balance exists between suitable PSA and LogP ranges that are compatible with promoting diffusion of small molecules across the blood-brain barrier while minimizing or avoiding potential metabolism by CYP2D6. Future investigations should be focused on further exploring the limited sphere of molecular property space that yields brain-penetrant, CYP2D6 stable compounds, to provide more appropriate guidance to CNS drug discovery campaigns.

Earlier models addressing small molecule CYP2D6 metabolism (Snyder et al., 2002; Haji-Momenian et al., 2003; Sciabola et al., 2007) have not been fully integrated into early stage drug discovery efforts partly because of limited accuracy in practical settings (Martin, 2006; Doweiko, 2008). Two potential factors affecting accuracy and utility are the limited number of data points and data quality used in model construction. For example, earlier quantitative structure-activity relationship studies (de Groot et al., 1997; Snyder et al., 2002; Haji-Momenian et al., 2003) used 40 to 50 compounds to build and validate models. In addition, most reports do not emphasize K_M values from purified enzyme sources, with one exception being the model reported by Snyder et al. (2002). The CYP2D6 kinetics database presented here addresses both of these issues by expanding the number of publicly available data points to 92 compounds and imposing inclusion and exclusion criteria for database entry. This expanded collection minimizes the confounding involvement of other enzymes in the K_M data.

It is important to note that one limitation with the database is the inclusion of only computed molecular property values. Deviations between experimental and computed values could lead to the generation of inaccurate or misleading patterns or trends. For example, differences between computed and experimental pK_a values of selected database members were observed (Table 1), suggesting a potential concern in using this particular property to forecast CYP2D6 substrates, despite previous reported correlations between pK_a and CYP2D6 substrates (Uphagrove and Nelson, 2001). In contrast, other computed properties were congruent with experimental values. Future expansion of the database must consider the quality and inclusion criteria for new data entries. As shown in the construction of the decision trees, the individual properties comprising the tree varied as a result of changes to the set of compounds used in construction. If the database becomes overpopulated with certain chemical families, the resulting patterns might deviate toward the properties most relevant to the high-abundance scaffolds. Notwithstanding, the database of disclosed CYP2D6 substrates and the reported statistical analyses of CYP2D6 substrates provide a useful tool for the future development

of new computational models, as well as a resource for individual drug discovery research campaigns with inclusion of future public disclosed data and internal proprietary data.

Acknowledgments. We thank Drs. Hantamalala Ralay Ranaivo and Linda J. Van Eldik for insightful comments and critical review of this manuscript.

References

- Arnaud-Neu F, Schwing-Weill MJ, Spiess B, and Yahya R (1990) Influence of substituents upon the basicity, spectral, characteristics, and lipophilicity of a series of 3-aminopyridazines. *J Chem Soc Perkin Transactions 2* **7**:1191–1196.
- Avdeef A, Comer JEA, and Thomson SJ (1993) pH-Metric Log P. 3. Glass electrode calibration in methanol-water, applied to pKa determination of water-insoluble substances. *Anal Chem* **65**:42–49.
- Clark DE (1999) Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. *J Pharm Sci* **88**:815–821.
- Davi H, Dupont P, Jeanniot JP, Roncucci R, and Cautreels W (1981) The biotransformation of [14C]minaprine in man and five animal species. *Xenobiotica* **11**:735–747.
- de Groot MJ, Bijloo GJ, Martens BJ, van Acker FA, and Vermeulen NP (1997) A refined substrate model for human cytochrome P450 2D6. *Chem Res Toxicol* **10**:41–48.
- Doweyko AM (2008) QSAR: dead or alive? *J Comput Aided Mol Des* **22**:81–89.
- Gleeson MP (2008) Generation of a set of simple, interpretable ADMET rules of thumb. *J Med Chem* **51**:817–834.
- Goodwin JT and Clark DE (2005) In silico predictions of blood-brain barrier penetration: considerations to “keep in mind.” *J Pharmacol Exp Ther* **315**:477–483.
- Guengerich FP, Miller GP, Hanna IH, Martin MV, Léger S, Black C, Chauret N, Silva JM, Trimble LA, Yergey JA, et al. (2002) Diversity in the oxidation of substrates by cytochrome P450 2D6: lack of an obligatory role of aspartate 301-substrate electrostatic bonding. *Biochemistry* **41**:11025–11034.
- Haji-Momenian S, Rieger JM, Macdonald TL, and Brown ML (2003) Comparative molecular field analysis and QSAR on substrates binding to cytochrome P450 2D6. *Bioorg Med Chem* **11**:5545–5554.
- Hu W, Ranaivo HR, Roy SM, Behanna HA, Wing LK, Munoz L, Guo L, Van Eldik LJ, and Watterson DM (2007) Development of a novel therapeutic suppressor of brain proinflammatory cytokine up-regulation that attenuates synaptic dysfunction and behavioral deficits. *Bioorg Med Chem Lett* **17**:414–418.
- Kalgutkar AS, Zhou S, Fahmi OA, and Taylor TJ (2003) Influence of lipophilicity on the interactions of *N*-alkyl-4-phenyl-1,2,3,6-tetrahydropyridines and their positively charged *N*-alkyl-4-phenylpyridinium metabolites with cytochrome P450 2D6. *Drug Metab Dispos* **31**:596–605.
- Kingsford C and Salzberg SL (2008) What are decision trees? *Nat Biotechnol* **26**:1011–1013.
- Kirchheiner J, Schmidt H, Tzvetkov M, Keulen JT, Lotsch J, Roots I, and Brockmoller J (2007) Pharmacokinetics of codeine and its metabolite morphine in ultra-rapid metabolizers due to CYP2D6 duplication. *Pharmacogenomics J* **7**:257–265.
- Kramer JA, Sagartz JE, and Morris DL (2007) The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. *Nat Rev Drug Discov* **6**:636–649.
- Leeson PD and Springthorpe B (2007) The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* **6**:881–890.
- Martin YM (2006) What works and what does not: lessons from experience in a pharmaceutical company. *QSAR Comb Sci* **25**:1192–1200.
- Obach RS and Reed-Hagen AE (2002) Measurement of Michaelis constants for cytochrome P450-mediated biotransformation reactions using a substrate depletion approach. *Drug Metab Dispos* **30**:831–837.
- Quinlan JR (1993) *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, CA.
- Sciabola S, Morao I, and de Groot MJ (2007) Pharmacophoric fingerprint method (TOPP) for 3D-QSAR modeling: application to CYP2D6 metabolic stability. *J Chem Inf Model* **47**:76–84.
- Snyder R, Sangar R, Wang J, and Ekins S (2002) Three-dimensional quantitative structure activity relationship for CYP2D6 substrates. *QSAR* **21**:357–368.
- Uphagrove AL and Nelson WL (2001) Importance of amine pKa and distribution coefficient in the metabolism of fluorinated propranolol derivatives. Preparation, identification of metabolite regioisomers, and metabolism by CYP2D6. *Drug Metab Dispos* **29**:1377–1388.
- Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, and Hassanali M (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res* **36**:D901–D906.
- Witten IH and Frank E (2005) *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Mateo, CA.
- Yu A, Dong H, Lang D, and Haining RL (2001) Characterization of dextromethorphan O- and N-demethylation catalyzed by highly purified recombinant human CYP2D6. *Drug Metab Dispos* **29**:1362–1365.

Address correspondence to: Laura K. Chico, 303 E. Chicago Avenue, Ward 8-196, Northwestern University, Chicago, IL 60611. E-mail: chico@u.northwestern.edu