

Predicting CYP2C19 catalytic parameters for enantioselective oxidations using artificial neural networks and a chirality code



Jessica H. Hartman^a, Steven D. Cothren^b, Sun-Ha Park^c, Chul-Ho Yun^c, Jerry A. Darsey^d, Grover P. Miller^{a,*}

^a Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, 4301 W. Markham, Slot 516, Little Rock, AR 72205, USA

^b Solibyte Solutions, Conway, AR 70234, USA

^c School of Biological Sciences and Technology, Chonnam National University, Gwangju 500-575, Republic of Korea

^d Department of Chemistry, University of Arkansas at Little Rock, Little Rock, AR 72204, USA

ARTICLE INFO

Article history:

Received 2 February 2013

Revised 3 April 2013

Accepted 12 April 2013

Available online 22 April 2013

Keywords:

Cytochrome P450

CYP2C19

Artificial neural networks

Catalytic parameters

Chirality codes

Enantioselective

Enantiomer

ABSTRACT

Cytochromes P450 (CYP for isoforms) play a central role in biological processes especially metabolism of chiral molecules; thus, development of computational methods to predict parameters for chiral reactions is important for advancing this field. In this study, we identified the most optimal artificial neural networks using conformation-independent chirality codes to predict CYP2C19 catalytic parameters for enantioselective reactions. Optimization of the neural networks required identifying the most suitable representation of structure among a diverse array of training substrates, normalizing distribution of the corresponding catalytic parameters (k_{cat} , K_m , and k_{cat}/K_m), and determining the best topology for networks to make predictions. Among different structural descriptors, the use of partial atomic charges according to the CHelpG scheme and inclusion of hydrogens yielded the most optimal artificial neural networks. Their training also required resolution of poorly distributed output catalytic parameters using a Box–Cox transformation. End point leave-one-out cross correlations of the best neural networks revealed that predictions for individual catalytic parameters (k_{cat} and K_m) were more consistent with experimental values than those for catalytic efficiency (k_{cat}/K_m). Lastly, neural networks predicted correctly enantioselectivity and comparable catalytic parameters measured in this study for previously uncharacterized CYP2C19 substrates, *R*- and *S*-propranolol. Taken together, these seminal computational studies for CYP2C19 are the first to predict all catalytic parameters for enantioselective reactions using artificial neural networks and thus provide a foundation for expanding the prediction of cytochrome P450 reactions to chiral drugs, pollutants, and other biologically active compounds.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Cytochromes P450 (CYP for specific enzymes) are potent catalysts that play critical roles in biological processes.^{1,2} These heme-containing enzymes catalyze many chemically challenging oxidative reactions such as hydroxylations of an unactivated carbon–hydrogen bond, dehalogenations, N- and O-dealkylations, epoxidations, and oxidations of nitrogen and sulfur.³ The molecular targets for these reactions include an array of structurally and functionally diverse compounds, many of which are chiral, such

as drugs, hormones, biosignaling molecules, dietary compounds (vitamins, nutraceuticals, and food additives), and pollutants. Unlike typical enzymes, cytochromes P450 often demonstrate broad substrate specificities, thereby making the prediction of whether a molecule binds and undergoes metabolism a significant challenge, especially during the early stages of drug development. Historically, a resolution to this problem involves the use of biochemical approaches to determine the metabolic efficiency of reactions between cytochromes P450 and their substrates. This biochemical strategy can require significant costs in time and resources. Computational methods lack these investments, and thus they have come to the fore as attractive alternatives for predicting the enzymatic properties of cytochromes P450.

During the last two decades, computational approaches have made significant progress toward understanding and possibly predicting interactions between cytochromes P450 and small molecules. In practice, these efforts have largely focused on predicting binding properties for molecules, reported as IC_{50} , K_i , K_d , and in

Abbreviations: CYP2C19, cytochrome P450 2C19; MDMA, 3,4-methylenedioxy-N-methylamphetamine; MDEA, 3,4-methylenedioxy-N-ethylamphetamine; MBDB, N-methyl-1,3-benzodioxolylbutanamine; BDB, 3,4-methylenedioxy- α -ethylphenethylamine; ANN, artificial neural network; HPLC, high performance liquid chromatography.

* Corresponding author. Tel.: +1 501 526 6486; fax: +1 501 686 8169.

E-mail address: MillerGroverP@uams.edu (Grover P. Miller).

limited cases, providing insights on the metabolism of substrates. Strategies for identifying these quantitative structure–activity relationships (QSAR) attempt to link the biological property of interest (output) to the energy-minimized structure of either the molecule alone or bound ('docked') in the active site of the enzyme (inputs).^{4,5} Due to the lack of known three-dimensional structures for cytochromes P450, the earliest QSAR methods involved multi-linear regression analyses that eventually gave way to more powerful machine languages employing non-linear regression analyses.^{6,7} For example, support vector machines have been used to predict whether molecules are substrates or inhibitors⁸ and even the K_m for substrates.⁹ Similarly, artificial neural networks can distinguish between substrates and inhibitors for multiple cytochromes P450^{10,11} and have been utilized to predict reaction rates (V_{max}) for certain N-dealkylations.¹² Another common ligand-based approach is comparative molecular field analysis (CoMFA). This strategy utilizes the common structural features among the molecules, including a class of enantioselective non-steroidal aromatase inhibitors,¹³ to predict their capacity to inhibit cytochrome P450 activity.^{14,15} Through CoMFA, it also is possible to distinguish inhibitors and substrates¹⁶ as well as predict K_m values.¹⁷ Despite these advances, most computational methods have not been adequately shown to predict catalytic parameters (k_{cat} , K_m , and k_{cat}/K_m) for the metabolism of these molecules and importantly, consider the impact of chirality on these processes.

Given biological systems have evolved in a chiral environment, the chirality of a molecule, such as a drug, often contributes significantly to biological function; accounting for molecular chirality is then critical for developing applications for computational algorithms. Chirality exists solely in three-dimensional space. In the absence of a chiral environment, the structures of the enantiomers have no energetic differences and are, thus, indistinguishable from one another. Consequently, the structure of these molecules cannot be linked to their biological properties by typical computational approaches. Overcoming this challenge requires formulation of a way to describe the unique chiral properties for a molecule. Several strategies generating chiral descriptors that distinguish between enantiomers were reviewed recently.^{18,19} Of these, Aires-de-Sousa and Gasteiger developed chirality codes representing the chirality of the molecule through a spectrum-like, fixed-length numerical trace that incorporates bond lengths, the molecular geometry, and property of atoms in neighborhoods about the chiral center. The use of these chirality codes by artificial neural networks and other regression models has predicted chromatographic enantioselectivity in high-performance liquid chromatography,²⁰ nuclear magnetic resonance chemical shifts in chiral solvents,²¹ and enantiomeric excess for reactions catalyzed by asymmetric non-enzymatic catalysts.^{22,23} To date, the potential power of these chirality codes has not been leveraged to predict enzymatic activity, especially for cytochromes P450, and importantly, the quantitative parameters describing the efficiency of enzyme catalysis.

Herein, we identified the most optimal artificial neural networks using conformation-independent chirality codes to predict CYP2C19 catalytic parameters (k_{cat} , K_m , and k_{cat}/K_m) for enantioselective reactions of known substrates. CYP2C19 metabolizes a structurally and functionally diverse array of chiral biological signals, dietary compounds, drugs, and pollutants such as pesticides.²⁴ This prominent role for CYP2C19 has led to the design of pharmacophore-based models to specifically inhibit its activity.²⁵ From the literature we culled available catalytic data for 23 chiral substrates analyzed under identical reaction conditions. This structurally diverse set of compounds includes 11 pairs of enantiomers plus a compound whose cognate enantiomer is not significantly metabolized by CYP2C19. For neural network

independent variables (inputs), we minimized three-dimensional structures for molecules with and without hydrogen atoms and calculated the partial atomic charges and then used the structural information to calculate conformation-independent chirality codes.^{7,26,27} For dependent variables (outputs), all three sets of catalytic parameters for CYP2C19 substrates were normalized to a GAUSSIAN distribution. Feed-forward artificial neural networks were trained with these inputs and outputs using the back propagation of errors algorithm and their topology optimized to retain predictability but avoid over-fitting of the data using the early stopping method.^{28,29} Performance of the artificial neural networks was assessed by leave-one-out cross correlation and then against previously uncharacterized CYP2C19 substrates. Specifically, we determined experimentally catalytic parameters for *R*- and *S*-propranolol oxidation by CYP2C19 and then compared them to those predicted by networks trained in this study. This combination of computational and biochemical approaches marks a significant step toward identifying optimal artificial neural network designs and validating their potential to predict catalytic parameters for chiral enzymatic reactions.

2. Materials and methods

2.1. The training set

The biological importance of CYP2C19 has led to a significant body of work exploring its catalytic properties; nevertheless, only a subset of that data is suitable for the computational studies explored in this study. We employed two criteria to select enantiomeric pairs of CYP2C19 substrates and their accompanying catalytic data for training artificial neural networks. Firstly, experimental conditions are known to influence catalytic parameters and thus we chose the largest data set of CYP2C19 substrates studied under identical experimental conditions. Secondly, implementation of conformation-independent chirality codes requires the presence of a single carbon chiral center,^{7,26,27} which added another limitation to identifying a suitable data set. Once the set was identified, the structures of these CYP2C19 substrates were used to generate chiral descriptors as independent input variables for the artificial neural networks, while the normalized catalytic parameters served as the dependent output variables.

2.2. Molecular modeling

The derivation of chiral descriptors for the CYP2C19 substrates required their partial atomic charges based on molecular modeling of the three-dimensional structures. In all cases, a convention was introduced whereby molecules with ionizable groups were modeled in the uncharged state. We then built the molecular set and optimized the structures initially using the MM:UFF method as implemented by ArgusLab 4.01 software (Arguslab Inc., Seattle, USA). Further minimization of the structures was accomplished using the B3LYP/6-31G* basis set with GAUSSIAN 03 version 6.0 software (Gaussian Inc., Wallingford, USA). The DFT/B3LYP functional was selected based on comparative studies between B3LYP, *ab initio*, and semi-empirical theories that demonstrated the B3LYP method led to better QSAR models when molecular geometries and energies were considered.³⁰ We calculated the corresponding partial atomic charges by Mulliken population analysis or based on electrostatic potentials using GAUSSIAN 03 software. Even though Mulliken population analysis is the default option for the software program, we included the option of electrostatic potentials using the grid-based CHelpG method for computing atomic charges, because this method is able to more

accurately model Coulomb potentials on the surface of the molecule.³¹

2.3. Conformation-independent chirality code as chiral descriptors (inputs)

The optimized geometries and partial atomic charges for CYP2C19 substrates were used to generate the conformation-independent chirality codes (CICC) for molecules^{7,26,27} as the independent input variables for the artificial neural networks. The chirality code transforms the various structures of the diverse set of substrates into a fixed-length code using a function derived from a radial distribution function. We designed in-house custom software written in C++ to calculate the chirality codes. Each molecule was divided into four neighborhoods, each belonging to atoms A, B, C, and D attached directly to the chiral center and every combination of four atoms *i*, *j*, *k*, and *l* considered for each neighborhood. The first component of this chirality code was E_{ijkl} , as defined by Eq. 1. In this equation, a_i represents the partial atomic charge of atom *i*, and r_{ij} is the distance between atoms *i* and *j*. The second major component of the chirality code is the chirality signal, S_{ijkl} , which takes on a value of +1 or –1.

The two values, E_{ijkl} and S_{ijkl} were calculated for all combinations of four atoms, each belonging to a different neighborhood and are combined to generate a chirality code using Eq. 2, where u was the running variable and B a smoothing factor. This function was then evaluated for a number of discrete values, in our case, 100 points, to obtain the same number of descriptors for each molecule. The range of the running variable u was determined empirically according to the range of atomic properties (atomic charges) in the data set. In practice, the range was set wide enough to include the features from all molecules in the set. The smoothing factor (B) controls the width of the peaks obtained by a graphical plot of $f_{CICC}(u)$ versus u .

$$E_{ijkl} = \frac{a_i a_j}{r_{ij}} + \frac{a_i a_k}{r_{ik}} + \frac{a_i a_l}{r_{il}} + \frac{a_i a_k}{r_{ik}} + \frac{a_j a_l}{r_{jl}} + \frac{a_k a_l}{r_{kl}} \quad (1)$$

$$F_{CICC}(u) = \sum_i \sum_j \sum_k \sum_l S_{ijkl} e^{-B(u-E_{ijkl})^2} \quad (2)$$

Two sets of codes for these molecules were generated in which hydrogen atoms not bonded directly to the chiral center were either considered or excluded. These possibilities were explored in this study based on their potential impact on the model.³² After calculating chirality codes for the CYP2C19 substrates, we then pared down the number of descriptors, and hence inputs for the neural networks, through a two-step strategy. Firstly, any points, which exhibited negligible (<0.00001) values for all molecules were excluded. Secondly, correlated inputs were excluded based on a Pearson correlation analysis. A global analysis of the Pearson correlation coefficients was performed for each descriptor in a search for two-descriptor correlations. The descriptors were then sorted with decreasing correlation coefficient and removed stepwise to avoid correlations greater than 0.85.

2.4. Normalization of catalytic parameters (outputs)

We are the first to employ all catalytic parameters for enzymatic reactions as the dependent output variables for artificial neural networks. Parameters for each compound obtained from the literature included the Michaelis constant K_m , enzymatic rate constant k_{cat} , and metabolic efficiency, that is, k_{cat}/K_m . In some cases, CYP2C19 demonstrated regioselectivity toward more than one site during the reaction, such as for enantiomers of bufuralol, limonene, methylbenzodioxolyl butanamine (MBDB), methylenedioxymethyl amphetamine (MDMA), methylenedioxyethyl amphetamine

(MDEA), phenprocoumon, and warfarin. If sites of metabolism were distal from one another, that is, greater than 5 Å, then only catalytic parameters for dominant metabolic pathway(s) were included in this study. For more proximal sites, we utilized apparent parameters reflecting the overall catalytic capacity toward the substrate (Appendix A, Table A.1). K_m Values for each metabolic pathway were averaged to yield $K_{m,ap}$. This approach is not likely to compromise the analysis, because the original values were always within error, which is reasonable given common binding modes for proximal sites of oxidation. The overall rate of substrate oxidation is critical for this study, and thus we summed up all rate constants to yield $k_{cat,tot}$ for network training purposes.

Neural networks learn more quickly and perform better if the data used to train the network is normally distributed. Consequently, we employed the Shapiro–Wilk test to assess the normality of the catalytic parameters for CYP2C19 substrates.³³ An initial analysis of untransformed data indicated poor normality, that is, non-Gaussian distribution of the data (see Section 3). For this reason, several methods of normalization were employed to improve the normality of the set, specifically, a square root operation, a logarithmic transformation, and a Box–Cox transformation.³⁴ Each normalized set of catalytic parameters was then subjected to the

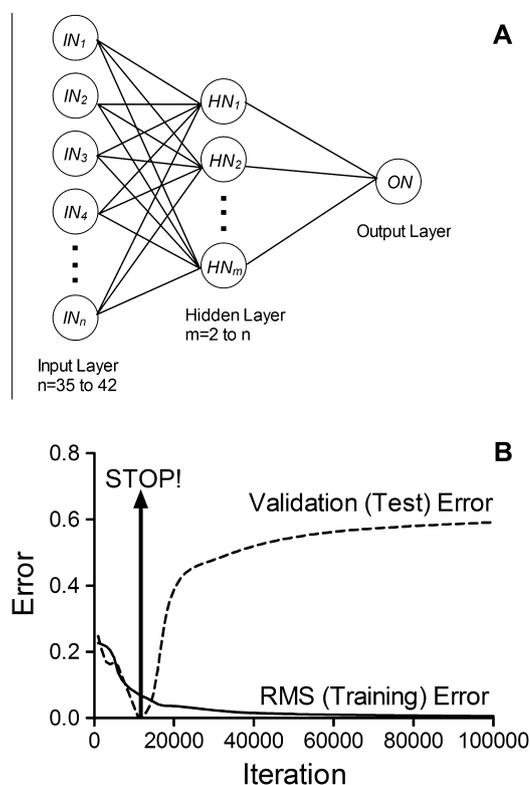


Figure 1. Topology and training of artificial neural networks. Structure of a fully connected three-layer feed-forward artificial neural network is shown in Panel A. The architecture included input nodes (INs) derived from the chirality codes whose final number (35–42) depended on the removal of all correlated inputs. The hidden nodes (HNs) were varied from 2 to *n* (the total number of input nodes) in order to achieve optimal performance. A single output node (ON) was used, representing the catalytic parameter for that particular training. Further details are provided in Section 2. An example of errors during a typical training is shown in Panel B. The solid line represents the root-mean-square (RMS) error on the training set. The dashed line represents the error on the validation compound, which was not included in the training set. The stopping point, as indicated on the plot, was then taken at the point before the validation error began to rise.

Shapiro–Wilk normality test to identify the best method of normalization.

2.5. Neural network model and structure

Feed-forward artificial neural networks trained these inputs and outputs using back propagation of errors algorithm as implemented by Nets software.³⁵ As shown in Figure 1, Panel A, the network architecture included an input layer consisting of selected chirality code descriptors, a hidden layer, and one output neuron (catalytic parameters i.e., $k_{\text{cat,tot}}$, $K_{\text{m,ap}}$, or $k_{\text{cat,tot}}/K_{\text{m,ap}}$). The most common method of hidden layer optimization is trial and error;²⁸ we optimized the number of hidden nodes by initially training networks with a large number of hidden nodes (number of hidden nodes = number of input nodes) and then repeated the training with a decreasing number of hidden nodes until performance of the network decreased to identify the minimal number of hidden nodes in which optimal network performance was retained as reported by others.

Networks were trained to convergence by the early stopping method to avoid over-fitting due to the relatively small number of training examples available for study.^{28,29} During the training process, the root mean square error of the training set and the error on the validation compound were saved as well as the weights at specified intervals. During the initial phase of training, the validation error would normally decrease and then increase when data began to be over-fitted. At that transition, weights and biases at the minimum of the validation error were used for network predictions. An example plot of errors from a typical training with the appropriate stopping point is shown in Figure 1, Panel B.

2.6. Model selection by end point leave-one-out cross validation analysis

An end point leave-one-out cross-validation was used to determine the most optimal chirality code configuration using electrostatic potentials as implemented by CHelpG or Mulliken Population analysis and including or excluding hydrogens not bonded to the chiral carbon. In this method, one compound is left out while the network is trained on the remaining compounds. The error on the compound left out is monitored and the stopping point determined as described previously. The error on the validation compound at the stopping point is therefore taken as the end-point cross-validated error.

Each of these neural network configurations underwent extensive trainings. Stochastic (on-line) gradient descent relies on updates to single weights performed randomly; therefore, multiple trainings can differ in their performance. The trainings were repeated until no gain in performance was obtained from additional trainings. The best-performing networks (with the smallest cross-validation error when training was stopped) were then used. To compare the models generated from chirality codes incorporating

Mulliken charges and those derived from the electrostatic potential, as well as to investigate the importance of the exclusion or inclusion of hydrogen atoms, a plot of the experimental values versus the neural network values for each trained network was examined. Specifically, the coefficient of determination of a linear least-squares regression analysis was used as a metric for network performance. The model with the highest coefficient of determination (R^2) was considered to be best and used for generating predictions. If models had similar coefficient of determinations, the model with the slope of the least squares regression line closest to one was considered to be best and used for prediction.

After visual identification of potential outlier points on the plots of the neural network values versus the experimental values, all networks were subjected to a second round of analysis whereby they were subjected to a ‘robust’ fit using GraphPad Prism 5. This type of analysis assumes that the variation around the curve/line follows a Lorentzian distribution rather than a Gaussian distribution, so unlike a least-squares regression it is more forgiving of potential outlier points. In both types of analyses; however, all points were considered (no data points were removed or ignored).

2.7. Predicting catalytic parameters for an uncharacterized substrate, propranolol

2.7.1. Strategy for validating the potential of networks

We assessed the power of optimally trained artificial neural networks to accurately predict CYP2C19 catalytic parameters for previously uncharacterized chiral test substrates, namely *R*- and *S*-propranolol. These drugs undergo oxidative *N*-dealkylation by CYP2C19 to corresponding enantiomers of norpropranolol.³⁶ Nevertheless, the catalytic parameters for this reaction have not been reported, hence we determined $k_{\text{cat,tot}}$, $K_{\text{m,ap}}$, and $k_{\text{cat,tot}}/K_{\text{m,ap}}$ for CYP2C19 using reaction conditions identical to those employed in the set of training compounds used for the neural networks. We then employed networks trained in this study to predict CYP2C19 catalytic parameters for *R*- and *S*-propranolol metabolism. The results from these experimental and computational efforts were then compared to identify the strengths and potential limitations of neural networks.

2.7.2. Catalytic parameters for *R*- and *S*-propranolol metabolism determined experimentally

We determined CYP2C19 steady-state catalytic parameters for *R*- and *S*-propranolol to norpropranolol. All chemicals were ACS grade or higher and obtained from Sigma–Aldrich. CYP2C19 supersomes containing oxidoreductase (but no cytochrome b_5) were obtained from BD Biosciences. For reactions, 50 nM CYP2C19 was incubated with eight concentrations of propranolol between 5.0 and 400 μM in 50 mM potassium phosphate buffer, pH 7.4, at 37 °C. Based on solubility, all propranolol stocks were prepared in methanol. Reactions contained a final concentration of 1% methanol, which has been reported to have little inhibitory effect.³⁷ Reactions were initiated upon addition of NADPH regenerating system (final concentration: 1 mM NADP⁺, 3 mM glucose-6-phosphate, 3 mM MgCl₂, and 1 U glucose-6-phosphate dehydrogenase) and quenched after 30 min with an equal volume of 0.4 N perchloric acid containing 50 μM internal standard (6-hydroxycoumachlor). Quenched reactions were clarified by centrifugation and supernatants recovered for subsequent HPLC analyses. Each set of reactions was performed at least in duplicate and replicated over multiple days.

We developed an in house HPLC method for the quantitation of norpropranolol. Samples containing *R*- or *S*-propranolol (substrate), norpropranolol (product) and internal standard (6-hydroxycoumachlor) were injected onto a Waters Breeze HPLC system

Table 1

Results of leave-one-out cross-correlation for networks trained with optimal artificial neural network architectures

Atomic charge method	Hydrogens considered ?	K_{m}		k_{cat}		$k_{\text{cat}}/K_{\text{m}}$	
		R^2	Slope	R^2	Slope	R^2	Slope
CHelpG	No hydrogen	0.58	1.1	0.61	1.0	0.27	1.0
	With hydrogen	0.75	0.98	0.67	1.2	0.40	0.92
Mulliken	No hydrogen	0.18	0.71	0.10	0.67	0.030	-0.37
	With hydrogen	0.35	0.93	0.81	1.2	0.34	0.82

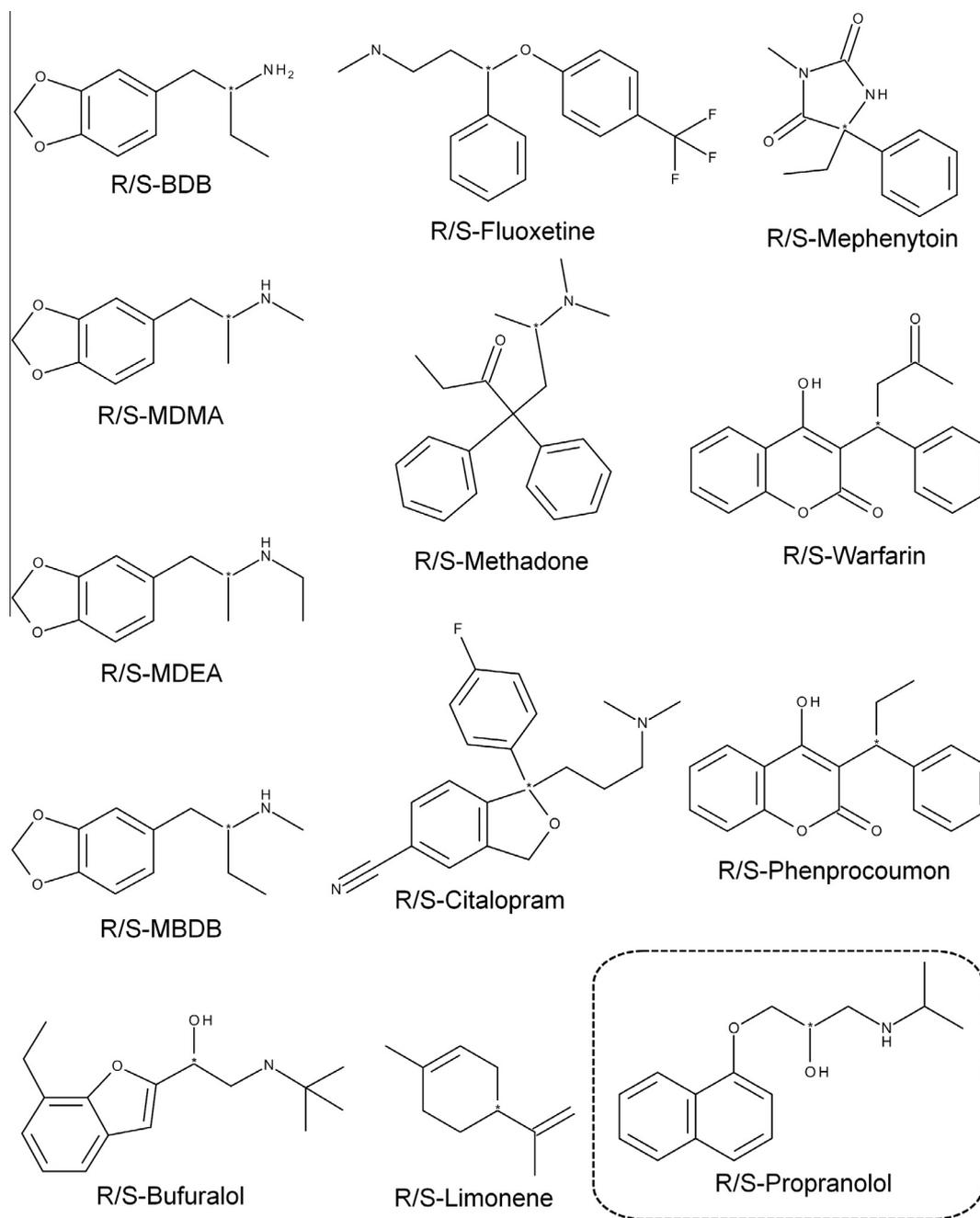


Figure 2. Two-dimensional structures of enantiomeric pairs of molecules included in the training set. Chirality centers are designated by an asterisk. Propranolol, indicated by the dashed box, was not included in the training set but instead was used as an independent test compound.

and resolved using a 4.6×150 mm Zorbax Eclipse $5 \mu\text{m}$ XDB-C18 column (Agilent) heated to 45°C using a gradient method involving $\text{H}_2\text{O}/0.1\%$ formic acid and methanol at 1.2 mL/min. Compound elution was monitored by fluorescence (excitation 230 nm; emission 380 nm) and peak areas normalized to internal standard for quantitation relative to known standards. Under these conditions the detection limit of norpropranolol was less than 16 nM (actual concentration in reaction).

Analysis of the kinetic data was performed by an iterative least-squares nonlinear regression analysis using GraphPad Prism 5 software (La Jolla, CA). The data were fit to the Michaelis–Menten equation for determination of the kinetic parameters $k_{\text{cat,tot}}$ and $K_{\text{m,ap}}$.

2.7.3. Catalytic parameters for *R*- and *S*-propranolol metabolism determined computationally

The generation of artificial neural network inputs for propranolol enantiomers was carried out as described in Sections 2 and 3. The two-dimensional structure of propranolol (shown in dashed box in Fig. 2) was modeled and the resulting optimized geometries and partial atomic charges used to generate the conformation-independent chirality codes (inputs) for the molecules.^{7,26,27} From these values, the same number of descriptors for *R*- and *S*-propranolol was used for propagation through the networks. The weights from each of the 23 leave-one-out cross-validated networks were used to generate predictions and confidence intervals for each of the three catalytic parameters.

2.7.4. Comparison of experimentally and computationally derived kinetic parameters

The experimentally derived kinetic parameters were compared to the computational predictions for the parameters. The predicted parameters were then divided by the experimental parameters and plotted on a log scale to show the relative deviation from experimental values and the spread of the data. In the case of an accurate prediction, the data would cluster at a value of zero on the *y*-axis. In this way, we could visualize the performance of the 23 networks in terms of overall prediction.

3. Results

3.1. Inputs

We found all available catalytic parameters for a structurally and functionally diverse array of 23 chiral substrates including pharmacologically active compounds (drugs and substances of

abuse), a solvent, and an insecticide, as shown in Figure 2 (parameters given in Appendix A, Table A.1).^{38–48} Molecules possessed various combinations of alkyl groups, aromatic rings, heteroaromatic rings, ether groups, secondary and tertiary amines, and carbonyl groups, yet shared a common single chiral carbon. Data for R-mephenytoin metabolism by CYP2C19 employed reaction conditions not selected for use in this study (*vide infra*),⁴⁹ and in fact, the authors reported a significantly low activity at a high substrate concentration indicating that the determination of catalytic parameters was not possible. Thus, the metabolism of R-mephenytoin could not be included in our study. Altogether, there were catalytic parameters for 23 molecules determined under identical reaction conditions, that is, 50 mM phosphate buffer pH 7.4 with identical enzyme preparation. All studies employed membrane fractions from insect cells co-expressing CYP2C19 with P450 reductase with no cytochrome *b*₅ present. These preparations were obtained commercially from BD/Gentest with the exception of one study on fluoxetine, which employed an in-house membrane

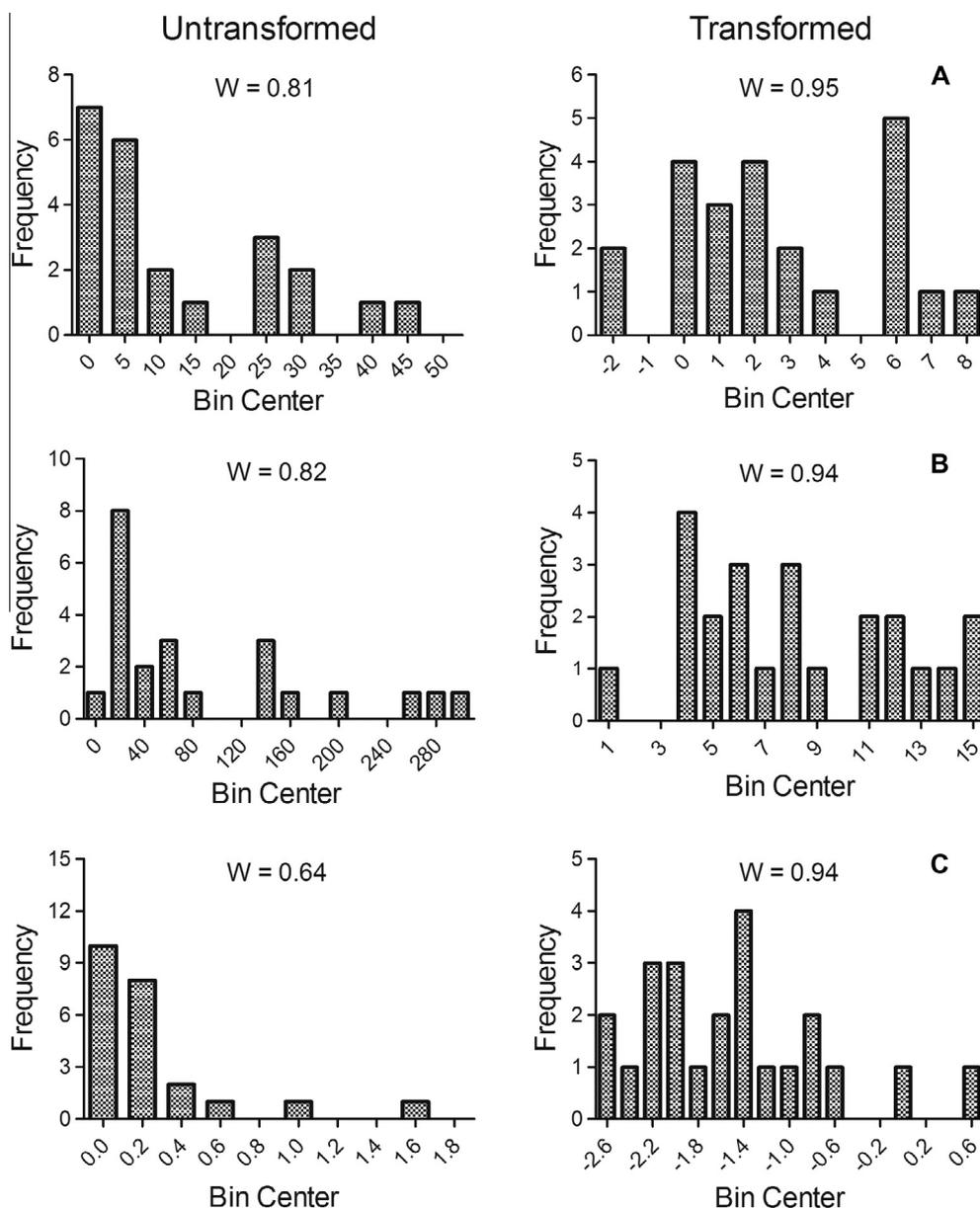


Figure 3. Normalization of catalytic parameters as data outputs. Histograms displaying distribution of data for untransformed and Box-Cox transformed catalytic parameters (dependent variable outputs) for $k_{cat,tot}$ (Panel A), $K_{m,ap}$ (Panel B), and $k_{cat,tot}/K_{m,ap}$ (Panel C), respectively.

preparation from insect cells co-expressing CYP2C19 with P450 reductase in the absence cytochrome b_5 .⁴¹ These compounds displayed a wide range in the catalytic efficiency and stereoselectivity of the reaction by CYP2C19. The parameters for these reactions are summarized in Appendix A, Table A.1.

The chirality code was determined from the structures of these molecules and evaluated at a range of the running variable 'u' (−0.16 to 0.16) sufficient to include all features for compounds tested. After evaluation of descriptor correlation, there were 35–42 descriptors remaining when the correlated descriptors were removed. This variance in the final number of remaining descriptors depended on the origination of the descriptors. The inclusion or exclusion of hydrogen atoms and the method of determining partial atomic charge impacted differently both the magnitude and number of total peaks in the function for the respective chirality codes.

3.2. Optimal normalization of network outputs by Box–Cox transformation

For artificial neural networks, especially with small training sets, normalization of a data set is very important, and thus we explored multiple normalization methods for the sets of $k_{\text{cat,tot}}$, $K_{\text{m,ap}}$, and $k_{\text{cat,tot}}/K_{\text{m,ap}}$ data corresponding to each of the 23 training compounds. All three sets of untransformed data initially failed the Shapiro–Wilk normality test ($p = 0.05$ cutoff) with W values of 0.81, 0.82, and 0.62 for $k_{\text{cat,tot}}$, $K_{\text{m,ap}}$, and $k_{\text{cat,tot}}/K_{\text{m,ap}}$, respectively, (Fig. 3). In particular, the set of data for catalytic efficiencies ($k_{\text{cat}}/K_{\text{m}}$) was skewed in favor of low efficiencies. The data were then treated independently with square root, logarithmic, and Box–Cox transformations. Although all techniques improved normality based on the Shapiro–Wilk test, the best results were obtained from the Box–Cox transformation yielding

W and lambda values of 0.95 and 0.32 for $k_{\text{cat,tot}}$, 0.94 and 0.33 for $K_{\text{m,ap}}$, and 0.94 and 0.30 for $k_{\text{cat,tot}}/K_{\text{m,ap}}$, respectively. When these values were used to perform the Box–Cox transformation, the normality of the data greatly improved for all three parameters (Fig. 3).

3.3. Identifying the best of neural network architecture

The optimal number of hidden nodes for the training of the networks was determined by initially using a large number of hidden nodes equal to the number of input nodes (35–42 total). The number of hidden nodes was then halved until a final number of two hidden nodes was reached. The performance of the networks was evaluated by careful examination of the cross-validation error, and the best performance was obtained when the number of hidden nodes was equal to one-half the number of input nodes (17–21 total). Training results are summarized in Appendix B, Tables B.1–5.

3.4. Determining the most optimal networks to predict catalytic parameters

The best method for computing the chirality codes to predict $k_{\text{cat,tot}}$, $K_{\text{m,ap}}$, and their ratio, $k_{\text{cat,tot}}/K_{\text{m,ap}}$, was the same for all three parameters as determined by leave-one-out cross-correlation analysis (Table 1). A comparison between the predicted parameter when network training was ceased and the experimental parameter is shown in Figure 4 for all three outputs. There was slightly better performance for $k_{\text{cat,tot}}$ obtained using Mulliken population analysis (GAUSSIAN 03 default); however, the best overall network performances employed the CHeLP method for computing partial atomic charges, and thus those networks were used for subsequent

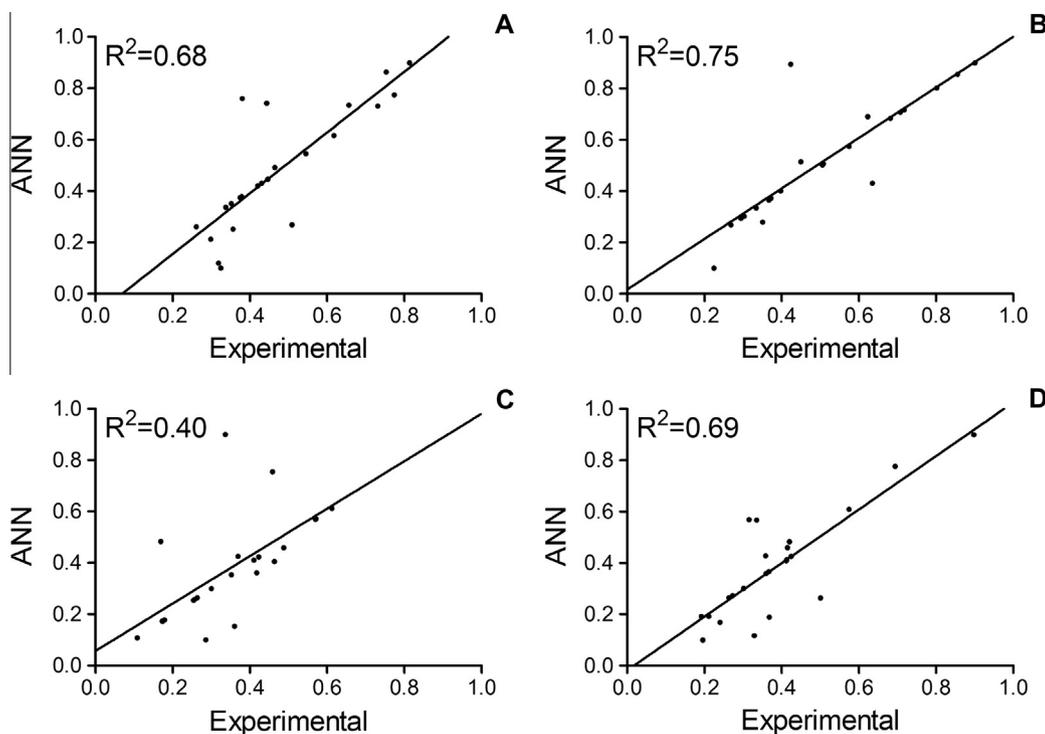


Figure 4. Leave-one-out end point cross-correlations for most optimal artificial neural networks. Plots reflect predicted values by artificial neural networks (ANN) values versus experimental values for best $k_{\text{cat,tot}}$ (Panel A), $K_{\text{m,ap}}$ (Panel B) and $k_{\text{cat,tot}}/K_{\text{m,ap}}$ (Panel C) networks. Panel D reflects the efficiencies ($k_{\text{cat,tot}}/K_{\text{m,ap}}$) for each compound derived from predictions of the individual parameters. Data shown represent catalytic parameters which have been subjected to Box–Cox transformation and linearly scaled between 0 and 1 (see Section 2.4 for more information on normalization). Data were fit to a least-squares linear regression using GraphPad Prism 5.0 and the coefficient of determination (R^2) is shown on the plots.

studies. In all cases, networks trained to predict parameters performed better when all hydrogens were considered, rather than only the one attached to the chiral carbon. Training results are included in Appendix B, Tables B.1–5.

Overall, networks trained to predict individual parameters produced better-performing networks than those trained to predict the ratio of the two. The data were analyzed by a traditional regression analysis as well as a robust regressive analysis, which is more

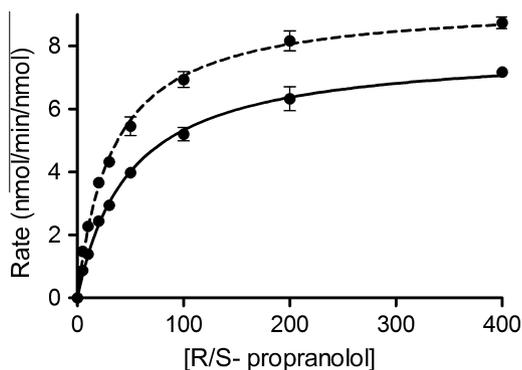


Figure 5. Steady-state oxidation of *R*- and *S*-propranolol by recombinant CYP2C19. Solid line represents *R*-propranolol; dashed line represents *S*-propranolol. Data were fit to the Michaelis–Menten equation using GraphPad Prism 5.0. For reactions, 50 nM recombinant CYP2C19 supersomes, propranolol (varied from 5 to 400 μ M), and a NADPH regenerating system were incubated at 37 °C and pH 7.4. The reported values represent the average from five individual experiments, including the mean \pm standard deviation.

Table 2

Experimentally determined parameters for *R*- and *S*-propranolol *N*-dealkylation by CYP2C19^a

Compound	k_{cat}^b	K_m^c	k_{cat}/K_m^d
<i>R</i> -Propranolol	7.9 (7.6–8.3)	49 (42–55)	0.16
<i>S</i> -Propranolol	9.4 (9.0–9.8)	33 (29–37)	0.28

^a 95% Confidence intervals are shown in parentheses.

^b Units are in nmol/min/nmol CYP2C19.

^c Units are in μ M.

^d Units are $\text{min}^{-1} \mu\text{M}^{-1}$.

forgiving of outlier data. The networks trained with $k_{cat,tot}$ displayed good performance with a small variability in the resulting performance of the networks. The R^2 obtained in the cross correlation analysis was 0.67 with a slope of 1.2. When the data were fit to the outlier-forgiving robust fit, the slope remained at 1.2 while the relative average deviation was 0.073. The best overall networks were those trained for $K_{m,ap}$ based on an R^2 of 0.75 and slope of 0.98. The results indicated relatively little variability in the predictions and a reasonable positive correlation. When the data were fit to a robust fit analysis, the slope was 1.0 and the relative average deviation was 0.0037. Finally, the performance of the networks trained with the catalytic efficiencies ($k_{cat,tot}/K_{m,ap}$) had more variability reflecting an R^2 of 0.40 and a slope of 0.92. This training benefited considerably from a robust fit analysis due to several outlier points, weighing heavily in the traditional least-squares regression. The slope for the robust fit for catalytic efficiency was 0.97 and the relative average deviation was 0.053.

3.5. Slightly enantioselective CYP2C19 metabolism of propranolol

CYP2C19 metabolized *R*- and *S*-propranolol into norpropranolol based on co-elution of authentic standards as reported by others.³⁶ The kinetic profile was hyperbolic for each enantiomer (Fig. 5) and thus was fit to the traditional Michaelis–Menten scheme to determine kinetic parameters as shown in Table 2. The experimentally-determined $K_{m,ap}$ for *R*-propranolol was slightly higher than that for the *S*-enantiomer. However, the $k_{cat,tot}$ (maximal reaction rate) was significantly higher for the *S*-enantiomer. This higher maximal rate and slightly lower $K_{m,ap}$ accounted for an overall higher efficiency ($k_{cat,tot}/K_{m,ap}$) for *S*-propranolol by CYP2C19 when compared to that for *R*-propranolol.

3.6. Prediction of CYP2C19 catalytic parameters for *R*- and *S*-propranolol by optimally trained neural networks

We assessed the power of the previously trained artificial neural networks to predict catalytic parameters for *R*- and *S*-propranolol and compared them to the experimentally derived values from this study. The leave-one-out cross correlation analysis generated 23 individual networks for each catalytic parameter, and thus each of those networks was used to predict parameters for *R*- and

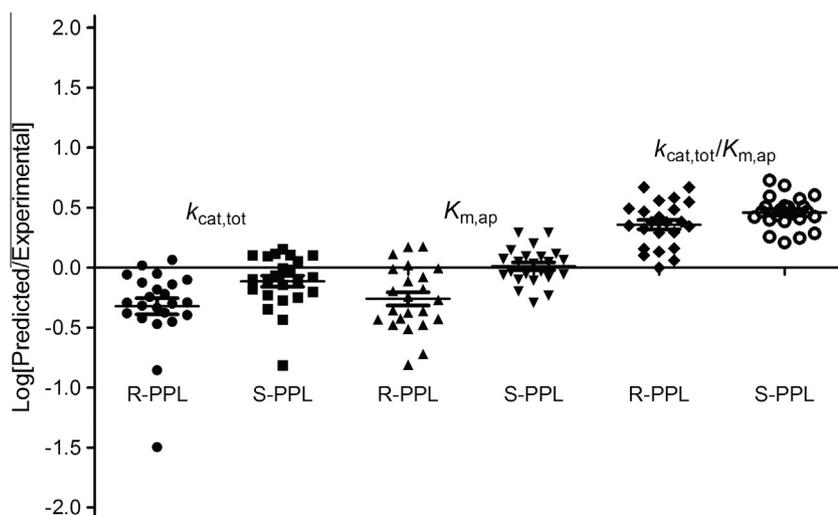


Figure 6. Predictions for *R*- and *S*-propranolol catalytic parameters by optimally trained artificial neural networks. Experimental/predicted catalytic parameters are shown in a log plot for Propranolol (PPL) predictions from all 23 networks using the best network configuration. The mean and 95% confidence intervals (shown by the solid bar and error bars on the column plot) for all parameters were used to generate a global prediction for that parameter.

Table 3
Computationally predicted parameters for R- and S-propranolol N-dealkylation by CYP2C19^a

Compound	k_{cat} ^b (predicted)	K_{m} ^c (predicted)	$k_{\text{cat}}/K_{\text{m}}$ ^d (predicted)	$k_{\text{cat}}/K_{\text{m}}$ ^{d,e} (calculated)
R-Propranolol	2.5 (2.0–3.0)	24 (17–30)	0.30 (0.25–0.36)	0.10
S-Propranolol	4.7 (3.8–5.5)	38 (33–44)	0.46 (0.40–0.52)	0.12

^a Predicted parameters reflect aggregate values derived from 23 neural networks as described in Section 3.6 under Results section. 95% Confidence intervals are shown in parentheses.

^b Units are in nmol/min/nmol CYP2C19.

^c Units are in μM .

^d Units are $\text{min}^{-1} \mu\text{M}^{-1}$.

^e Catalytic efficiencies calculated from parameters predicted individually by the respective artificial neural networks.

S-propranolol. The distribution of those predictions would validate the strategy for generating the respective networks and assess the potential for any strong data points biasing the training and hence predictions. Figure 6 displays the comparisons between the predicted and experimental values as a scatter plot. The y-axis is a log scale such that predictions matching the experimental values would result in a value of zero. Relative catalytic parameters for both substrates clustered near zero. In other words, networks made similar predictions for parameters and hence the exclusion of one data set from the leave-one-out analysis did not significantly compromise the ability of the networks to generalize predictions for test compounds.

Aggregate catalytic parameters ($k_{\text{cat,tot}}$, $K_{\text{m,ap}}$, and $k_{\text{cat,tot}}/K_{\text{m,ap}}$) predicted by the 23 respective neural networks are summarized in Table 3. The neural networks correctly predicted the enantioselectivity of the oxidative reaction, although the experimental $k_{\text{cat,tot}}$ values were 2- to 3-fold higher. The magnitude of the predicted $K_{\text{m,ap}}$ values were closer to those determined experimentally. The higher affinity of CYP2C19 toward S-propranolol observed in those experiments was not capitulated by the predicted $K_{\text{m,ap}}$ values, although CYP2C19 enantioselectivity on substrate binding was minimal. The predicted catalytic efficiencies for the reaction ($k_{\text{cat,tot}}/K_{\text{m,ap}}$) were about twofold higher in magnitude than those observed experimentally. Interestingly, the calculated catalytic efficiencies based on the predictions of the individual enantiomers were more consistent with experimental results both in magnitude and enantioselectivity toward S-propranolol.

4. Discussion

We are the first to successfully implement artificial neural networks employing chirality codes to predict all catalytic parameters for enantioselective CYP2C19 reactions including those involving many drugs. Optimization of the neural networks required identifying the most suitable representation of structure among a diverse array of training compounds, normalizing the distribution of the corresponding catalytic parameters (k_{cat} , K_{m} , and $k_{\text{cat}}/K_{\text{m}}$) for their metabolism, and determining the best topology for the networks to make predictions. An end point leave-one-out cross correlation of the best artificial neural networks revealed that the predictions for individual catalytic parameters (k_{cat} and K_{m}) were more consistent with experimental values than those for catalytic efficiency ($k_{\text{cat}}/K_{\text{m}}$) possibly owing to differences in the origination of the respective catalytic parameters. Lastly, the neural networks predicted correctly the enantioselectivity and comparable catalytic parameters to those measured in this study for previously uncharacterized CYP2C19 substrates, R- and S-propranolol. Taken together, these efforts support the practical application of these artificial neural networks to predict enantioselective reactions by CYP2C19.

4.1. Determining suitable structural descriptors as data inputs

Like many cytochrome P450 enzymes,^{1,2} CYP2C19 metabolizes a structurally and functionally diverse array of substrates including

the training set of enantiomers used in our study,²⁴ and thus it was necessary to accurately capture the salient features of the molecules for training the artificial neural networks. We chose the conformation-independent chirality code to reflect the chirality and structure of the molecule without limiting the structure to a particular conformation. These codes require calculations of the partial atomic charge for the molecule. The default GAUSSIAN 03 software method for this information is Mulliken population analysis. While this method is one of the oldest and simplest, it is also highly dependent on the choice of basis set and thus prone to error.⁵⁰ As an alternative, we employed the grid-based CHelpG method for computing atomic charges, because this method is able to more accurately model Coulomb potentials on the surface of the molecule.⁵¹ Moreover, the performance of some networks but not others was optimal when all hydrogens were included in the structure. The importance of hydrogens not bonded to the chiral carbon may reflect contributions through substructures, such as functional groups, to the properties of the molecule. In our case, the most optimal artificial neural networks resulted from the use of partial atomic charges according to the CHelpG scheme and the inclusion of hydrogens. These findings underscore the need to effectively model electrostatic potentials for the entire molecule mediating contacts to CYP2C19 during catalysis.

4.2. Novelty of predicting catalytic parameters for enantioselective reactions as data outputs

Catalytic parameters describe the mechanism by which enzymes drive chemical reactions and predict their behavior, such as within biological systems. Binding is the first step in the catalytic cycle, and historically those interactions have been amenable to prediction by a wide range of computational methods. A majority of QSAR studies, especially those used for drug development, have focused on the inhibitory potential of molecules,^{8,10,14,15} including enantiomers,¹³ toward metabolism by multiple cytochrome P450 isoforms. Nevertheless, binding does not necessarily result in the appropriate orientation of substrate to undergo catalysis and thus may not correlate with catalysis. Consequently, groups have employed various computational methods that differentiate between potential cytochrome P450 substrates and inhibitors as well as substrate K_{m} values^{9,17,52} and V_{max} for N-dealkylations.¹² Chirality has largely been ignored in computational methods for cytochrome P450 reactions. In this study we report a computational method to predict catalytic parameters for CYP2C19 enantioselective reactions of known substrates and thus advance the application of this strategy for biologically relevant reactions catalyzed by cytochromes P450. We did not include the ability to distinguish between substrate or inhibitor due to the presence of multiple reports on achieving that goal.⁵³

The distribution of CYP2C19 catalytic parameters was not ideal for training artificial neural networks. These networks achieve optimal performance when training values reflect a normal or Gaussian distribution. In our case, data were clustered at the low end of the range in values and hence significantly non-Gaussian.

This shortcoming may reflect the lack of sufficient data due to the relatively small sample size used in this study, that is, 23 training compounds. Although more data may resolve the issue of data distribution, we solved the problem in this case by identifying the Box–Cox transformation as the most optimal method to significantly improve the Gaussian distribution of data for successfully training artificial neural networks.

4.3. Implementing artificial neural networks to utilize inherent structural diversity of substrates for predictions

We sought a flexible and robust computational approach capable of exploiting the diverse set of enantiomeric training compounds to predict CYP2C19 catalysis. These molecules were mostly of chiral drugs and one dietary compound that possessed various combinations of structural and functional groups sharing a single chiral carbon. These training set characteristics can pose a challenge for computational QSAR studies that rely on common structural features to identify structure–function relationships, such as CoMFA.^{15,17} For our set of compounds, these types of computational approaches would force the exclusion of certain compounds from our diverse array of training set, and thus decrease critical structure–function information on CYP2C19 catalysis. Artificial neural networks do not require training compounds to share structural elements and are able to distinguish between substrates and inhibitors for multiple cytochromes P450^{10,11} and predict reaction rates (V_{\max}) for certain N-dealkylations.¹² For our study, the use of artificial neural networks provided the opportunity to utilize a diverse array of training compounds that more accurately reflected the catalytic capacity of CYP2C19 toward enantiomeric substrates.

The optimization of artificial neural network topology and training is not trivial, and is best addressed through a systematic approach. For this reason, we automated many steps in the training and analysis of networks. A script was written to run batch trainings, which ensured consistent training parameters over multiple runs and allowed monitoring of the errors generated to determine the best stopping point and avoid over-fitting as described in Section 2 and cited by others.²⁹ Upon analysis of networks trained with varying number of hidden nodes, it became clear that the performance of the network suffered with too many or too few hidden nodes. Too few hidden nodes caused the network functions to be too simple for representing the relationship between inputs and outputs; too many hidden nodes provided too many degrees of freedom, so that the network over-trained prematurely. Through our testing strategy, we were able to effectively identify the most optimal network topologies for predicting catalytic parameters for an enzyme for the first time.

4.4. Validation of trained artificial neural networks for predicting catalytic parameters for enantioselective reactions

The power of neural networks was assessed through end point leave-one-out cross-correlation that compared the internal performance of the network among the compounds in the training set. For this analysis, there were good correlations between predicted and experimental values for $k_{\text{cat,tot}}$ and $K_{\text{m,ap}}$. The slightly better performance with $K_{\text{m,ap}}$ is consistent with previous computational experiments by others^{9,17,52} and likely reflects the suitability of computational approaches to predict simple binding interactions. Under rapid equilibrium conditions K_{m} is essentially a binding constant K_{d} when the catalytic step is slow. The formation of the substrate–enzyme complex depends mainly surface properties including polarity, hydrophobicity, etc.

By contrast, the $k_{\text{cat,tot}}$ parameter is a more complex parameter than $K_{\text{m,ap}}$ and reflects the rate of substrate oxidation as well as the

contributions of other processes such as inactivation of the enzyme, alternate oxidative pathways and the uncoupling of the oxidative reaction altogether.⁵⁴ The contributions of these processes depend on substrate structure and its mode of interaction with CYP2C19 and collectively contribute to the apparent rate of product formation ($k_{\text{cat,tot}}$). Ideally, trained neural networks take into account all productive and nonproductive processes contributing to $k_{\text{cat,tot}}$, even if contributions of those processes are not identified specifically. For example, if substrate structure lends itself to enzyme inactivation, then neural networks would predict a lower rate of product formation than that for simple oxidation of substrate, which may apply to fluoxetine metabolism and inactivation of CYP2C19⁵⁵ and other substrates containing a methylenedioxy group. Similarly, the role of substrate structure favoring alternate reaction pathways and uncoupling of oxidation would result in a less efficient rate of product formation. Those aspects of the catalytic cycle would be taken into account by neural networks trained on known CYP2C19 reactions. Nevertheless, the effectiveness of this training depends on sufficient representation of those substrates prone to pathways other than substrate oxidation. It is not possible to determine whether this is the case for our study or not, such that the interpretation of the relationship between $k_{\text{cat,tot}}$ and structure is not obvious, even if the neural networks predict the relationship.

Although there were relatively good correlations for each individual catalytic parameter, the predicted and the experimental catalytic efficiencies, that is, the ratio of the parameters, for the reactions were not well correlated by this analysis. This outcome may reflect the different roles of structural properties contributing to the respective parameters that are not reconciled when training networks to catalytic efficiencies. Studies with β -glucosidase demonstrated that K_{m} correlated with hydrophobicity while specific activity correlated with dipole moment.⁵⁶ For our study, we resolved the challenge of predicting catalytic efficiencies by demonstrating that the ratio of individually predicted $k_{\text{cat,tot}}$ and $K_{\text{m,ap}}$ values correlated well with actual catalytic efficiencies for CYP2C19 enantioselective reactions.

While CYP2C19 metabolizes *R*- and *S*-propranolol,³⁶ the parameters for the reaction have not been reported previously. Propranolol is structurally related to bufuralol, another nonselective beta blocker drug that was included in the training set and had the highest metabolic efficiencies. Despite structural similarities between the drugs, *R*- and *S*-propranolol undergo hydroxylation of the isopropyl group of the amine and subsequent N-dealkylation, while bufuralol undergoes hydroxylation of the ethyl group side chain on the aromatic ring. Bufuralol cannot undergo N-dealkylation due to the presence of the *tert*-butyl group on the amine. Nonetheless, the enantioselectivity observed for propranolol metabolism was similar to that for bufuralol based on our experimental studies reported here. The metabolism of enantiomers of both drugs resulted in similar K_{m} values and a higher k_{cat} for the *S*-enantiomer. Overall, the catalytic efficiency toward oxidation of propranolol was less than bufuralol, but in the mid- to high range for the whole data set.

A practical application of the neural networks was to test their ability to predict our experimental CYP2C19 catalytic parameters for metabolism of *R*- and *S*-propranolol. The clustered distribution of values predicted by networks further validated the design of the networks to generate consistent predictions and demonstrated the absence of any single data points biasing predictions. Overall, networks correctly predicted the enantioselectivity of CYP2C19 reactions, akin to efforts by others toward the enantiomeric excess (or more appropriately the enantiomeric ratio, er)⁵⁷ for reactions.^{58,59} Importantly, our computational efforts also yielded critical and practical catalytic parameters describing the efficiency of the reaction as a function of concentration and time. Specifically,

neural networks predicted lower $k_{cat,tot}$ values than measured experimentally, but similar predicted and experimental values were observed for K_m . These trends mirrored findings from the leave-one-out cross correlations in that networks were better able to identify structure–function relationships involved in the recognition of substrates than their subsequent oxidation in the reaction. Nevertheless, these artificial neural networks can easily be re-trained with the incorporation of more data from CYP2C19 reactions, such as the results from reactions carried out in this study, and thus improve the ability of networks to accurately predict catalytic parameters for chiral CYP2C19 reactions.

5. Concluding remarks

Cytochromes P450 play a central role in biological processes especially those involving the metabolism of chiral drugs, and thus our development of a computational method for predicting quantitative parameters for enantioselective reactions by CYP2C19 marks an important advancement in the field of drug development. The successes in this study reflect a critical combination of chirality codes and the flexibility of artificial neural networks. The predictive power of these neural networks would likely improve through retraining with an expanded training set. This feat is possible given that many reactions were excluded in this study due to differences in experimental conditions that could impact the corresponding parameters. In following, our studies with R- and S-propranolol are a step in the right direction for providing data to make ‘smarter’ neural networks. That effort also highlighted the critical need for computational methods to transition from predicting known reactions to those that are unknown. Taken together, the findings in our study provide a template for realizing the true potential of computational approaches, such as chirality codes and neural networks, to advance their use for predicting biologically relevant chiral biological processes such as drug metabolism.

Acknowledgments

The authors are grateful to Dr. Joao Aires-de-Sousa (Universidade Nova de Lisboa) for helpful discussions on implementation of chirality codes and Dr. Ralph L. Kodell (University of Arkansas for Medical Sciences) for advice on methods for normalizing data. In addition, the authors are thankful for funding that supported this research. Grover P. Miller was supported by a bridging Grant from the University of Arkansas for Medical Sciences and a pilot study Grant provided by the National Institute of Health Grant No. UL1 TR000039 awarded to the University of Arkansas for Medical Sciences. This work was also partially supported by equipment purchased under NASA Award NCC5-597 and by the National Science Foundation under Grant CRI CNS-0855248, Grant EPS-0701890, Grant EPS-0918970, Grant MRI CNS-0619069, and OISE-0729792.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.bmc.2013.04.044>.

References and notes

- Bernhardt, R. J. *Biotechnol.* **2006**, *124*, 128.
- Guengerich, F. P. *Chem. Res. Toxicol.* **2008**, *21*, 70.
- Isin, E. M.; Guengerich, F. P. *Biochim. Biophys. Acta* **2007**, *1770*, 314.
- Sridhar, J.; Liu, J.; Foroozesh, M.; Stevens, C. L. *Molecules* **2012**, *17*, 9283.
- de Groot, M. J. *Drug Discovery Today* **2006**, *11*, 601.
- Lewis, D. F. V. *Toxicology* **2000**, *144*, 197.
- Gasteiger, J. J. *Med. Chem.* **2006**, *49*, 6429.
- Vasanthanathan, P.; Taboureau, O.; Oostenbrink, C.; Vermeulen, N.; Olsen, L.; Jorgensen, F. *Drug Metab. Dispos.* **2009**, *37*, 658.
- Leong, M. K.; Chen, Y.-M.; Chen, T.-H. *J. Comput. Chem.* **1899**, *2009*, 30.
- Molnar, L.; Keseru, G. M. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 419.
- Korolev, D.; Balakin, K. V.; Nikolsky, Y.; Kirillov, E.; Ivanenkov, Y. A.; Savchuk, N. P.; Ivashchenko, A. A.; Nikolskaya, T. *J. Med. Chem.* **2003**, *46*, 3631.
- Balakin, K. V.; Ekins, S.; Bugrim, A.; Ivanenkov, Y. A.; Korolev, D.; Nikolsky, Y. V.; Ivashchenko, A. A.; Savchuk, N. P.; Nikolskaya, T. *Drug Metab. Dispos.* **2004**, *32*, 1111.
- Cavalli, A.; Bisi, A.; Bertucci, C.; Rosini, C.; Paluszczak, A.; Gobbi, S.; Giorgio, E.; Rampa, A.; Belluti, F.; Piazzi, L.; Valenti, P.; Hartmann, R. W.; Recanatini, M. *J. Med. Chem.* **2005**, *48*, 7282.
- Rao, S.; Aoyama, R.; Schrag, M.; Trager, W. F.; Rettie, A.; Jones, J. P. *J. Med. Chem.* **2000**, *43*, 2789.
- Sridhar, J.; Foroozesh, M.; Stevens, C. L. *SAR QSAR Environ. Res.* **2011**, *22*, 681.
- Kemp, C.; Flanagan, J. U.; van Eldik, A. J.; Maréchal, J. D.; Wolf, C. R.; Roberts, G. C.; Paine, M. J.; Sutcliffe, M. J. *J. Med. Chem.* **2004**, *47*, 5340.
- Haji-Momenian, S.; Rieger, J. M.; Macdonald, T. L.; Brown, M. L. *Bioorg. Med. Chem.* **2003**, *11*, 5545.
- Del Rio, A. *J. Sep. Sci.* **2009**, *32*, 1566.
- Zhang, Q.-Y.; Xu, L.-Z.; Li, J.-Y.; Zhang, D.-D.; Long, H.-L.; Leng, J.-Y.; Xu, L. *J. Chemom.* **2011**, *26*, 497.
- Aires-de-Sousa, J.; Gasteiger, J. *J. Mol. Graphics* **2002**, *20*, 373.
- Zhang, Q.; Carrera, G.; Gomes, M.; Aires-de-Sousa, J. *J. Org. Chem.* **2005**, *70*, 2120.
- Aires-de-Sousa, J.; Gasteiger, J. *Comb. Chem.* **2005**, *7*, 298.
- Zhang, Q.; Zhang, D.; Li, J.; Zhou, Y.; Xu, L. *Chemom. Intell. Lab.* **2011**, *109*, 113.
- Rokitta, D.; Fuhr, U. *Curr. Drug Metab.* **2010**, *11*, 153.
- Foti, R. S.; Rock, D. A.; Han, X.; Flowers, R. A.; Wienkers, L. C.; Wahlstrom, J. L. *J. Med. Chem.* **2012**, *55*, 1205.
- Aires-de-Sousa, J.; Gasteiger, J. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 369.
- Aires-de-Sousa, J.; Gasteiger, J.; Gutman, I.; Vidović, D. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 831.
- Almeida, J. S. *Curr. Opin. Biotechnol.* **2002**, *13*, 72.
- Al-Jumaily, A.; Chen, L. *J. Theor. Biol.* **2012**, *310*, 115.
- Yan, X.; Xiao, H.; Gong, X.; Ju, X. *J. Mol. Struct.* **2006**, *764*, 141.
- Young, D. *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*; Wiley-Interscience: New York, 2001.
- Gasteiger, J.; Engel, T. *Cheminformatics*; Wiley-VCH: Weinheim, 2007; p 420.
- Shapiro, S.; Wilk, M. *Biometrika* **1965**, *52*, 591.
- Box, G.; Cox, D. *J. R. Stat. Soc. Ser. B* **1964**, *26*, 211.
- Zurada, J. M. *Introduction to Artificial Neural Systems*; West Publishing Company: New York, 1992.
- Yoshimoto, K.; Echizen, H.; Chiba, K.; Tani, M.; Ishizaki, T. *Br. J. Clin. Pharmacol.* **1995**, *39*, 421.
- Busby, W. F., Jr.; Ackerman, J. M.; Crespi, C. L. *Drug Metab. Dispos.* **1999**, *27*, 246.
- Meyer, M. R.; Peters, F. T.; Maurer, H. H. *Toxicol. Lett.* **2009**, *190*, 54.
- Narimatsu, S.; Takemi, C.; Tsuzuki, D.; Kataoka, H.; Yamamoto, S.; Shimada, N.; Suzuki, S.; Satoh, T.; Meyer, U. A.; Gonzalez, F. J. *J. Pharmacol. Exp. Ther.* **2002**, *303*, 172.
- Olesen, O. V.; Linnet, K. *Pharmacology* **1999**, *59*, 298.
- Margolis, J. M.; O'Donnell, J. P.; Mankowski, D. C.; Ekins, S.; Obach, R. S. *Drug Metab. Dispos.* **2000**, *28*, 1187.
- Miyazawa, M.; Shindo, M.; Shimada, T. *Drug Metab. Dispos.* **2002**, *30*, 602.
- Meyer, M. R.; Peters, F. T.; Maurer, H. H. *Biochem. Pharmacol.* **2009**, *77*, 1725.
- Meyer, M. R.; Peters, F. T.; Maurer, H. H. *Drug Metab. Dispos.* **2009**, *37*, 1152.
- Venkatakrishnan, K.; von Moltke, L. L.; Greenblatt, D. J. *J. Pharm. Sci.* **1998**, *87*, 845.
- Gerber, J. G.; Rhodes, R. J.; Gal, J. *Chirality* **2004**, *16*, 36.
- Ufer, M.; Svensson, J.; Krausz, K.; Gelboin, H.; Rane, A.; Tybring, G. *Eur. J. Clin. Pharmacol.* **2004**, *60*, 173.
- Kim, S.; Kang, J. Y.; Hartman, J. H.; Park, S. H.; Jones, D. R.; Yun, C. H.; Boysen, G.; Miller, G. P. *Drug Metab. Lett.* **2013**, *6*, 157.
- Goldstein, J. A.; Faletto, M. B.; Romkes-Sparks, M.; Sullivan, T.; Kitareewan, S.; Raucy, J. L.; Lasker, J. M.; Ghanayem, B. I. *Biochemistry* **1994**, *33*, 1743.
- Reed, A. E.; Weinstock, R.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*.
- de Molletta, F.; Angelotti, W.; Romero, R.; Montanari, C.; da Silva, A. *J. Mol. Model.* **2008**, *14*, 975.
- Wang, Y.-H.; Li, Y.; Li, Y.-H.; Yang, S.-L.; Yang, L. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4076.
- Zhang, T.; Dai, H.; Liu, L. A.; Lewis, D. F. V.; Wei, D. *Mol. Inform.* **2012**, *31*, 53.
- Guengerich, F. *Biol. Chem.* **2002**, *383*, 1553.
- Stresser, D. M.; Mason, A. K.; Perloff, E. S.; Ho, T.; Crespi, C. L.; Dandeneau, A. A.; Morgan, L.; Dehal, S. S. *Drug Metab. Dispos.* **2009**, *37*, 695.
- Yoon, M.; Nam, Y. K.; Choi, W. Y.; Sung, N. D. *J. Microbiol. Biotechnol.* **2007**, *17*, 1789.
- Gawley, R. E. *J. Org. Chem.* **2006**, *71*, 2411.
- Mazurek, S.; Ward, T. R.; Novic, M. *Mol. Divers.* **2007**, *11*, 141.
- Funar-Timofei, S.; Suzuki, T.; Paier, J. A.; Steinreiber, A.; Faber, K.; Fabian, W. M. *F. J. Chem. Inf. Comput. Sci.* **2003**, *43*, 934.