DMD Fast Forward. Published on May 30, 2007 as doi:10.1124/dmd.106.014613




# Computational approaches that predict metabolic intermediate complex formation with CYP3A4 (+$b_5$)

David R. Jones, Sean Ekins, Lang Li and Stephen D. Hall

Indiana University School of Medicine, Department of Medicine, Division of Clinical Pharmacology,Wishard Memorial Hospital, Myers Bldg W7123, Indianapolis, IN 46220. (D.R.J., S.D.H.)

ACT LLC, 1 Penn Plaza-36th Floor, New York, NY 10119 and Department of Pharmaceutical Sciences, University of Maryland, 20 Penn Street, Baltimore, MD 21201. (S.E.)

Indiana University School of Medicine, Department of Medicine, Division of Biostatistics, Wishard Memorial Hospital, Indianapolis, IN 46220 (L.L.)

DMD Fast Forward. Published on May 30, 2007 as DOI: 10.1124/dmd.106.014613
This article has not been copyedited and formatted. The final version may differ from this version.



**Running Title: Metabolic intermediate complex formation with CYP3A4 (+$b_5$)**

Stephen D. Hall, Ph.D.

Indiana University School of Medicine, Department of Medicine,

Division of Clinical Pharmacology, Wishard Memorial Hospital,

Myers Bldg W7123, Indianapolis, IN 46220

E-mail: sdhall@iupui.edu

Phone: 317-630-8795

Fax: 317-630-8185

| | |
|---|---|
| Text pages: | 19 |
| Tables: | 5 |
| Supplemental Tables : | 1 |
| Figures: | 5 |
| Supplemental Figures : | 3 |
| References: | 45 |
| Words in Abstract: | 243 |
| Words in Introduction: | 756 |
| Words in Discussion: | 1875 |

Abbreviations; MIC, metabolic intermediate complex; QSAR, quantitative structure activity relationship; RP, recursive partitioning

DMD Fast Forward. Published on May 30, 2007 as DOI: 10.1124/dmd.106.014613
This article has not been copyedited and formatted. The final version may differ from this version.



## ABSTRACT

Some mechanism-based inhibitors cause irreversible inhibition by forming a metabolic intermediate complex (MIC) with the cytochrome P450 (CYP). In the present study, 54 molecules (substrates of CYP3A and amine-containing compounds that are not known substrates of CYP3A) were spectrophotometrically assessed for their propensity to cause MIC formation with recombinant CYP3A4 (+$b_5$). Comparisons of common physicochemical properties showed that mean (± s.d.) molecular weight (MW) of MIC forming compounds was significantly greater than mean MW of non-MIC forming compounds, 472 (± 173) vs. 307 (± 137), respectively. Computational pharmacophores, logistic regression and recursive partitioning (RP) approaches were applied to predict MIC formation from molecular structure and to generate a quantitative structure activity relationship (QSAR). A pharmacophore built with SKF-525A, erythromycin, amprenavir, and norverapamil indicated that four hydrophobic features and a hydrogen bond acceptor were important for these MIC forming compounds. Two different RP methods using either simple descriptors or 2D augmented atom descriptors indicated that hydrophobic and hydrogen bond acceptor features were required for MIC formation. Both of these RP methods correctly predicted the MIC formation status with CYP3A4 for 10 out of 12 literature molecules in an independent test set. Logistic multiple regression and a third classification tree model predicted 11 out of 12 molecules correctly. Both models possessed a hydrogen bond acceptor and represent an approach for predicting CYP3A4 MIC formation that can be improved using more data and molecular descriptors. The preliminary pharmacophores provide structural insights that complement those for CYP3A4 inhibitors and substrates.

The cytochrome P450 (CYP) enzymes (EC 1.14.14.1) are membrane-bound proteins that catalyze many oxidations of hydrophobic endobiotics and xenobiotics. The catalytic activity of CYPs may be reduced by reversible and irreversible inhibition upon administration of xenobiotics. One type of irreversible inhibition, mechanism-based inhibition, has been the focus of many studies related to CYPs recently due to its clinical implications for prediction of drug-drug interactions. A mechanism-based inhibitor is one that binds to the active site, then becomes catalytically activated by the enzyme (Silverman, 1988). The activated form of the molecule will irreversibly bind to the enzyme to remove it from the active enzyme pool. Some mechanism-based inhibitors cause this irreversible inhibition by forming a metabolic intermediate complex (MIC) with the heme of the CYP (Franklin, 1977). Primary, secondary or tertiary amines, or methylenedioxy constituents (Supplemental Figure 1), in the molecule are prerequisites for compounds that chelate the heme of the CYP (Franklin, 1977). More recent studies with metabolites of molecules such as indinavir and nelfinavir that lack these functional groups yet still display MIC formation may indicate other chemical moieties are also involved (Ernest et al., 2005).

The CYP3A family of enzymes is recognized as perhaps the most important for human drug metabolism as they metabolize the majority of commercially available drugs (Wrighton et al., 2000). These CYPs are expressed in numerous tissues but affect xenobiotic metabolism and clearance mainly in the liver and small intestine. There are four differentially regulated CYP3A genes in humans, CYP3A4, CYP3A5, CYP3A7 and CYP3A43, of which, CYP3A4 is the most abundant form in the adult liver (Wrighton et al., 2000). The CYP3A forms demonstrate regioselectivity differences for some biotransformations of the same compounds, while CYP3A5 generally has lower or comparable metabolic capability than CYP3A4 for common probe substrates (Williams et al., 2002).

In early pharmaceutical drug discovery the assessment of inhibitory potency with CYP3A4 for new chemical entities is often included as a first tier in vitro screen using fluorescent probes (Crespi and Stresser, 2000). This type of screen is useful for identifying potential inhibition of co-

administered drugs and is commonly followed by time dependent inhibition studies (Wu et al., 2003). These studies may be supplemented with additional in vitro assays with more traditional drug substrate probes using LC/MS analysis (Ekins et al., 2000b). Data from all of these in vitro screens are increasingly applied to predictive algorithm development (Ekins et al., 2003a; Ekins et al., 2003b; Gao et al., 2002; Riley et al., 2001). The interest in computational models based on *in vitro* data for predicting potential drug interactions via this protein and others (Ekins and Swaan, 2004) represents a possible means to improve productivity of the drug discovery process and remove potential bottlenecks caused by in vitro testing. Due to their highly parallel nature computational methods are also likely the fastest and most cost-effective method for indication of likely toxic consequences (Ekins et al., 2000a), and suggesting new hypotheses for testing in vitro. A recent review of computational methods for CYPs has documented how these approaches have been used over nearly twenty years alongside empirical methods (de Graaf et al., 2005). Others have described and compared the many pharmacophores that have been generated for CYPs (de Groot and Ekins, 2002; Ekins et al., 2001) providing insight into the important features for interaction of ligands and the proteins. Computational pharmacophores for CYP3A4 have therefore been derived for substrates and inhibitors using kinetic constants $K_m$, $K_{i\ (apparent)}$ and $IC_{50}$ data. Several studies have shown differential MIC formation for compounds between CYP3A4 and CYP3A5. For example CYP3A5 did not form a MIC with verapamil (Wang et al., 2004) or saquinivir (Ernest et al., 2005). These differences suggest that the binding sites (and catalytic rate) for both enzymes are subtly different accommodating some molecules and not others as has been suggested by recent pharmacophores for inhibitors of both enzymes (Ekins et al., 2003b).

In the present study 54 molecules have been assessed for their propensity to form MIC while 27 molecules possess $k_{inact}$ data with recombinant CYP3A4 in vitro. These data have been analyzed along with the generation of simple molecular descriptors to understand any possible relationships between MIC, $k_{inact}$ and molecule structure. Several computational approaches namely, pharmacophores,

quantitative structure activity relationships (QSAR) (Ekins and Swaan, 2004), classification tree, and multiple regression methods were used to generate models to predict these properties separately. These models were applied to the prediction of the probability of a molecule forming an MIC with CYP3A4 using a series of test molecules, while the $k_{inact}$ models were internally validated by omitting molecules at random.

## MATERIALS AND METHODS

*Chemicals.* All of the following classes of study drugs were purchased from Sigma-Aldrich, United States Pharmacopeia, kindly donated by companies or by investigators. NADPH was purchased from Roche Diagnostics (Indianapolis, IN). All other chemicals were at least analytical grade.

Antibiotics: triacetyloleandomycin, erythromycin, N-desmethylerythromycin (United States Pharmacopeia, Rockville MD), clarithromycin (donated by Abbott Laboratories, Chicago, IL), N-desmethylclarithromycin (donated by Abbott Laboratories, Chicago, IL), 14-hydroxyclarithromycin (donated by Abbott Laboratories, Chicago, IL).

Calcium channel blockers: amlodipine, diltiazem, N-desmethyldiltiazem (donated by Tanabe Seiyaku Co, Osaka, Japan), desacetyldiltiazem (donated by Tanabe Seiyaku Co, Osaka, Japan), desacetyl-N-desmethyldiltiazem (donated by Tanabe Seiyaku Co, Osaka, Japan), R-verapamil, S-verapamil, norverapamil, D-617, nicardipine.

CNS drugs: amitriptyline, d,l-amphetamine, benzphetamine, brompheniramine, chlorpheniramine, desipramine, diphenhydramine, fenfluramine, fluoxetine, fluvoxamine, imipramine, loperamide, meperidine, methamphetamine, methylenedioxymethamphetamine, methylenedioxyethylamphetamine, 2-methylamino-1-(3,4-methylenedioxyphenyl)butane (MBDB), methylphenidate, mirtazepine, nefazodone, norfluoxetine, nortriptyline, orphenhydramine, paroxetine, phencyclidine, propoxyphene (United States Pharmacopeia, Rockville MD), sertraline, tranylcypramine.

HIV protease inhibitors: amprenavir (donated by GlaxoSmithKline, Research Triangle Park, NC), indinavir (donated by Merck, Whitehouse Station, NJ), nelfinavir (donated by Agouron Pharmaceuticals, Inc., New York, NY), ritonavir (donated by Abbott Laboratories, Chicago, IL), saquinavir (donated by F. Hoffmann-La Roche, Nutley, NJ), lopinavir (donated by Abbott Laboratories, Chicago, IL).

Anticancer drugs: tamoxifen, N-desmethyltamoxifen (donated by Zeneca Pharmaceuticals, Wilmington, DE), 4-hydroxytamoxifen, 3-hydroxytamoxifen.

Miscellaneous drugs: mifepristone, SKF-525A.

*Enzyme Preparation.* Insect cell membranes containing baculovirus cDNA-expressed CYP3A4 (+$b_5$) were purchased from BD Gentest™ (Bedford, MA). The CYP content was provided by the manufacturer at the time of purchase.

*Estimation of Mechanism-based Inactivation Parameters.* Testosterone 6β-hydroxylation was determined to quantify time- and concentration-dependent loss of CYP3A4 activity in the presence of inactivator (Ernest et al., 2005). The concentration of 6β-hydroxytestosterone was determined by HPLC with UV detection as described earlier (Zhao et al., 2002). The mechanism-based inactivation parameters, $k_{inact}$ and $K_I$, were obtained from the pseudo first order decline in the percentage of remaining CYP3A4 activity after preincubation with inactivator by nonlinear regression without weighting using WinNonlin Professional, v 4.0 (Pharsight, Mountain View, CA). Details of the data fitting and parameter estimation have been described previously (Ernest et al., 2005).

*Metabolic Intermediate Complex Measurements.* MIC formation was identified using dual wavelength spectroscopy (Uvikon 933 double beam UV/VIS Spectrophotometer; Research Instruments International, San Diego, CA) by scanning from 380 to 500 nm. The sample cuvette contained protein (200 pmol of CYP3A4 (+$b_5$)), 100 mM sodium phosphate buffer (pH = 7.4), inhibitor, and 1 mM NADPH (made with the phosphate buffer), whereas the reference cuvette contained protein, 100 mM phosphate buffer, vehicle, and 1 mM NADPH. All MIC formation experiments were initiated by the addition of NADPH and maintained at 37°C. The absorbance difference spectra for the identification of MIC formation were estimated by subtracting the absorbance at 490 nm of the absorbance scan from the difference of the absorbance scan at 60 min and

DMD Fast Forward. Published on May 30, 2007 as DOI: 10.1124/dmd.106.014613
This article has not been copyedited and formatted. The final version may differ from this version.



a background absorbance scan. MIC formation was quantified from absorbance difference spectra using an extinction coefficient of 65 $mM^{-1}$ $cm^{-1}$ (Pershing and Franklin, 1982).

*Computational modeling.* An initial assessment was undertaken to determine whether MIC and non-MIC forming compounds could be differentiated with simple calculated molecular properties. Calculated logP (octanol-water partition coefficient) and molecular weight were determined for each compound with ChemDraw for Excel (CambridgeSoft, MA). Statistical parameters were calculated with SPSS (v 12.0, SPSS Inc., Chicago, IL). Additionally, several different computational techniques were evaluated; a pharmacophore method (Catalyst, Accelrys, San Diego, CA), three recursive partitioning (tree) methods including ChemTree, (Golden Helix Inc., Bozeman, MT), Cerius$^2$, (Accelrys, San Diego, CA) and tree function in R software v2.2.1 (www.r-project.org), using different types of descriptors; a linear model (multiple regression and regression tree model) using R software, v 2.2.1 and a logistic model using R software, v 2.2.1.

*Pharmacophores for MIC prediction.* The Catalyst software was used on a Silicon Graphics Octane workstation. After importing the molecular structures, conformers were generated for each compound using the BEST functionality for each molecule and limited to a maximum of 255 conformers with an energy range of 20 kcal/mol. Structural relationships between compounds that form MIC and those that do not form MIC were assessed separately using the common features function (HIPHOP). The following groups of molecules were selected in order to test the fit of the hypothesis with unknowns: Compounds that form MIC (SKF-525A, erythromycin, amprenavir, norverapamil), compounds that do not form MIC (mifepristone, sertraline, 4-hydroxytamoxifen, and paroxetine), and compounds that do not form MIC but inactivate CYP3A4 (lopinavir, saquinavir, nefazodone). These alignments were then used to fit molecules to the hypothesis.

*Recursive partitioning (ChemTree) for MIC prediction.* The ChemTree recursive partitioning software was run on a Pentium 4 processor. The 54 molecules and experimental data (binary response; 1= compound that forms MIC, 0 = compound that does not form detectable MIC) were imported as a

DMD Fast Forward. Published on May 30, 2007 as DOI: 10.1124/dmd.106.014613
This article has not been copyedited and formatted. The final version may differ from this version.



.sdf file in ChemTree to generate over 330 path length descriptors (Young et al., 2002). These descriptors were used to generate either 1 tree or 100 random tree models (with the following options; p value threshold for splits, 0.99; max segments, 3; parallel threads, 1; and resampling iterations, 10,000).

*Recursive partitioning (Cerius$^2$ CSAR) for MIC prediction*. The Cerius$^2$ 4.8 software was used to generate the following default descriptors: sum of atomic polarizabilities, dipole magnitude, radius of gyration, area, molecular weight (MW), molecular volume, density, principal moment of inertia, rotatable bonds, hydrogen bond acceptors, hydrogen bond donors, and AlogP98 (a second method for calculating the octanol-water partition coefficient). The CSAR recursive partitioning method (Hawkins et al., 1997) was used with the 54 molecule training set (with the following settings, using weighted observations equally, gini scoring, scaled prune (0)). This model was internally validated using cross-validation: e.g. 10-fold, five-fold, or two-fold. A 10-fold cross-validation leaves out 10% test data.

*Recursive partitioning (tree function in R) for MIC prediction.* The R v2.2.1 software was used to predict MIC formation based on the twelve prescribed descriptors. The recursive partitioning method (Breiman et al., 1984) was used with the 54 molecule training set (with weighted observations equally and gini scoring). This model was internally validated using five-fold cross-validation, and the tree is pruned based on misclassification.

*Logistic Regression for MIC prediction*. Logistic regression model was implemented in a R function, generalized linearized model, glm() (R v 2.2.1; www.r-project.org). The model was generated with the twelve Cerius$^2$ descriptors to predict MIC formation. Fifty four compounds served as the training set. The glm() algorithm is based on the maximizing likelihood approach. A forward step-wise forward variable selection strategy was used to select descriptors. The optimal model was internally validated based on a five-fold cross-validation.

Pharmacophore, tree models, and the logistic regression model were tested with molecules outside of the training set using recently published data for twelve compounds (Chatterjee and

Franklin, 2003; Kajita et al., 2002; Kasahara et al., 2000; Kim et al., 2001; Wu et al., 2003; Yamazaki and Shimada, 1998) with MIC formation for CYP3A4 (Table 3).

*Linear Multiple Regression and Regression Tree Models for $k_{inact}$, $K_I$, and $k_{inact}/K_I$ prediction.* The $k_{inact}$ and $K_I$ values were estimated from in vitro studies with 27 compounds as described previously. Linear multiple regression and regression tree models were fit with the twelve descriptors, listed in the recursive partitioning section above, to predict $k_{inact}$, $K_I$, $k_{inact}/K_I$. Stratified randomization was utilized to divide samples into three strata based on 33 percentile and 67 percentile of $k_{inact}$ and/or $K_I$ values. Within each stratum, 2/3 of the samples (n=18) were randomly selected into the training set and the remaining 1/3 (n=9) were selected into the external validation set. All variables were log-transformed.

Two R functions, linear model, lm(), and tree() (R v 2.2.1), were implemented for regression tree model construction. These two predictive models were based on five-fold cross-validation on training set samples.

**RESULTS**

*Formation of Metabolic Intermediate Complex***.** MIC formation with each compound and CYP3A4 (+$b_5$) was determined spectrophotometrically. Figure 1 illustrates absorbance difference changes by SKF-525A, a compound that inactivates CYP3A4 and forms MIC, by phencyclidine, a compound that is not metabolized by CYP3A4 and does not form MIC, and lopinavir, a compound that inactivates CYP3A4 but does not form MIC. The minimum detectable absorbance difference is 0.002, which corresponds to 31 pmol of complex or 16 % of total CYP3A4. The percentage of maximum MIC was estimated by comparing the amount of MIC that was formed with the amount of CYP3A4 (+$b_5$) that was used in the experiment. If a compound formed less than 16 % of maximum MIC then it was considered a compound that does not form MIC. The MIC formation results are shown in Table 2.

*Mechanism-based Inactivation Parameters.* The $k_{inact}$ and $K_I$ values for 27 compounds were estimated from in vitro preincubations with CYP3A4 (+$b_5$). The estimates for each compound are listed in Table 1. The $k_{inact}$ values ranged from 0.03-1.12 $min^{-1}$ and $K_I$ values ranged from 0.1-12.5 μM.

*Computational Methods.* An initial comparison of the distributions of the molecular weights of 27 molecules that formed MIC to the 27 molecules that do not form MIC (Table 2, Supplemental Figure 2) indicated mean values of 472 (standard deviation = 174, range 263.4 - 798) and 308 (standard deviation = 137, range 133.2 – 670.9), respectively, which were statistically significant applying a *t*-test with $p<0.05$. The mean calculated log P for MIC forming compounds, 3.92 (standard deviation = 1.44, range 1.44-6.81) was not significantly different when compared with the mean calculated log P for non-MIC forming compounds, 3.86 (standard deviation = 1.53, range 1.24-7.05), applying the same statistical tests.

*Pharmacophores for MIC prediction*. Catalyst was used to generate common feature alignment (HIPHOP) pharmacophores for compounds that formed MIC and those that did not. These preliminary

pharmacophores represent the disposition of common chemical features such as hydrogen bond acceptor, hydrogen bond donor and hydrophobe in 3D space. The pharmacophore based on the four MIC forming compounds (SKF-525A, erythromycin, amprenavir, and norverapamil) indicated that four hydrophobic features and a hydrogen bond acceptor were common (Figure 2a). The pharmacophore for non-MIC forming compounds was smaller and possessed two hydrophobes and one hydrogen bond donor (Figure 2b). The pharmacophore for compounds that were CYP3A4 inactivators but did not form MIC was considerably larger and more feature rich (5 hydrophobes and 2 hydrogen bond acceptor features) than the MIC pharmacophore (Figure 2c). These pharmacophores represent a qualitative approach and the MIC forming compound pharmacophore was used to evaluate the mapping of several other literature compounds including (+) beta-hydrastine, dimethyl-4,4'-dimethoxy-5,6,5',6'-dimethylenedioxybiphenyl-2,2'dicarboxylate (DDB) and roxithromycin which show a good agreement with most of the features (Figure 3).

*Recursive partitioning (ChemTree) for MIC prediction.* Two different recursive partitioning methods were used to produce models for MIC prediction. The first method was used to produce a single tree model and the second model produced a 100 random tree model by applying ChemTree to all 54 molecules in the training set. The single tree model (Figure 4) was produced that used two path length descriptors (Young et al., 2002), C(CCC)-C(O) (e.g. this descriptor represents a carbon atom attached to three other carbon atoms at a distance from a carbon-oxygen bond) and C(CCC)-N(CC), out of 330 general descriptors that were initially generated. This model misclassified seven molecules in the training set based on a score less than 0.5 for compounds that do not form MIC and a score greater than 0.5 for compounds that do form MIC, such that 87 % were correctly predicted (Table 2). The 100 random tree model generated with the same training set used 21 out of the 330 available descriptors. The most frequently used descriptors were C(CCC)-N(CC), C(CCC)-C(CCC), C(CCC)-C(CN), C(N)-C(O) and C(CC)-N(CC) (representing connectivity between the atoms listed). When the

observed versus predicted values for the 100 tree model were analyzed, five of the 54 compounds (91 % accuracy) were misclassified as non-MIC forming compounds (Table 2).

The single tree and 100 tree models both incorrectly assigned two of the twelve compounds in the validation test set (Supplemental Figure 3) representative of 83 % accuracy (Table 3). These compounds were olopatadine M1 and miocamycin.

*Recursive partitioning (*Cerius$^2$ CSAR*) for MIC prediction*. The second recursive partitioning tree model was generated with twelve descriptors in Cerius$^2$ (Figure 5 and Supplemental Table 1 available online only). This CSAR model classified correctly all of the molecules in the training set using the same 54 compounds that were utilized in the ChemTree models. This model was internally validated using 10-fold, 5-fold, and 2-fold cross validation, resulting in 74 %, 80 % and 76 % correct predictions, respectively. This stability in the predictions for the left out molecules perhaps also suggests the stable nature of this model for external testing.

This model incorrectly assigned roxithromycin M2 and M3 as non MIC forming and miocamycin as MIC forming, three of twelve compounds, in the validation test set, which represents 75 % accuracy (Table 3).

The combination of Cerius$^2$ molecular descriptors with ChemTree descriptors enabled another 100 tree model to be built with ChemTree. This used 37 of 342 available descriptors. The most frequently used descriptors in addition to those path length descriptors previously mentioned, included the radius of gyration from the Cerius$^2$ descriptors. This 100 tree ChemTree recursive partitioning model was used to generate predictions for molecules that were not in the training set and the mean score determined as described above. This model was used to score the validation set of literature molecules for MIC formation and incorrectly assigned olopatidine M1 and miocamycin as MIC forming, two of twelve which represents 83 % accuracy (Table 3).

*Recursive partitioning (*tree function in R*) for MIC prediction*. Another recursive partitioning tree model (logistic model tree function in R) was generated using the twelve descriptors from Cerius$^2$

(Figure 6 and Supplemental Table 1 available online only). This tree model was formed by optimizing the gini function. It was then pruned by minimizing the misclassification rate in a five-fold cross-validation. The final tree model for the 54 training set molecules classified correctly 96 % of the molecules in the training set and contained hydrogen bond acceptor, radius of gyration, sum of atomic polarization, and dipole magnitude descriptors. This model was also used to predict the tendency of the literature test set molecules to form a MIC. Among twelve compounds in the external validation set, the model incorrectly assigned miocamycin as MIC forming, which relates to 92 % accuracy (Table 3).

*Logistic Regression Model for MIC prediction.* Logistic regression was generated with a R function generalized linear model (glm) and the twelve Cerius$^2$ molecular descriptors (supplemental Table 1) to predict MIC formation. Logistic regression utilizes a logit transformation of descriptors in linear combinations to predict the probability of a binary outcome variable. Usually, maximum likelihood estimation is utilized for estimating regression parameters. Fifty four drug compounds served as the training set (Table 2), and twelve compounds represented the external validation set (Table 3). The step-wise feed forward procedure was chosen to select and optimize prediction model. The internal validation was based on a five-fold cross-validation.

The logistic regression training data produced a predictive model in which the hydrogen bond acceptor descriptor was the most important predictor (p-value = 0.002). The following equation was used to predict MIC formation applying the hydrogen bond acceptor descriptor.

$$\Pr(MIC = +) = \frac{e^{-2.47+0.59\times \text{Hbond\_Acceptor}}}{1+e^{-2.47+0.59\times \text{Hbond\_Acceptor}}}$$

The logistic regression model had 80 % prediction accuracy for the 54 compounds in the training set. Among 12 compounds in the external validation set, it incorrectly assigned miocamycin as MIC forming, which represents 92 % prediction accuracy (Table 3).

*Regression Tree Models for $k_{inact}$, $K_I$, and $k_{inact}/K_I$ prediction.* Recursive partitioning tree models (tree function in R) were generated with twelve descriptors in Cerius$^2$ (Table 1 and

DMD Fast Forward. Published on May 30, 2007 as DOI: 10.1124/dmd.106.014613
This article has not been copyedited and formatted. The final version may differ from this version.



supplemental table 1) for $k_{inact}$, $K_I$, and $k_{inact}/K_I$ prediction. The tree models were formed by optimizing the gini function. It was then pruned by minimizing the sum of square of error in a five-fold cross-validation. The final tree models exhibited coefficients of determination ($R^2$) of 0.41, 0.39, 0.64, for $k_{inact}$, $K_I$, and $k_{inact}/K_I$ prediction, respectively, in the eighteen training samples and $R^2$ values of 0.12, 0.15, 0.05 in the nine validation samples (Table 4). The $k_{inact}/K_I$ model contained sum of atomic polarizabilities and hydrogen bond donor descriptors; the $k_{inact}$ model contained AlogP98 and hydrogen bond donor, and the $K_I$ possessed dipole magnitude and density descriptors.

*Linear Regression Models for $k_{inact}$, $K_I$, and $k_{inact}/K_I$ prediction.* Linear regression was implemented with a R function *lm*. A step-wise feed forward procedure and a five-fold cross validation were used to select optimal prediction models. The final regression models provided $R^2$ values of 0.16, 0.26, and 0.40 in the eighteen training samples for $k_{inact}$, $K_I$, and $k_{inact}/K_I$, prediction respectively, and $R^2$ values of 0.23, 0.26, and 0.11 in the nine validation samples (Table 4). The $k_{inact}/K_I$ model contained the rotatable bond descriptor; the $k_{inact}$ model contained the radius of gyration descriptor, and the $K_I$ model possessed the sum of atomic polarization descriptor.

## DISCUSSION

Considerable research has focused on computational algorithms that can be used in drug discovery for predicting molecular properties. These include rule based models for predicting the likelihood of absorption (Lipinski et al., 1997), methods that have more graphical outputs for predicting binding to CYPs (Ekins et al., 2001) or abstract QSAR methods applied to many proteins and properties relevant to ADME/Tox (Ekins and Swaan, 2004) . Such calculations can be performed with very large numbers of molecules to act as a molecule selection filter. Comparative molecular fields analysis (CoMFA) and pharmacophore approaches have been used to model CYP enzymes involved in drug metabolism and which have a role in important drug-drug interactions (de Groot et al., 1999a; de Groot et al., 1999b; de Groot et al., 1996; Ekins et al., 2001; Jones et al., 1996) . Recursive partitioning methods have been used extensively with large sets of molecules and either continuous (Chen et al., 1999; Chen et al., 1998) or binary data, for therapeutic target end points as well as CYP inhibition (Ekins et al., 2003a) and toxicity properties such as AMES mutagenicity status (Young et al., 2002). Linear regression is perhaps one of the earliest approaches used for QSAR of P450 (Hansch and Zhang, 1993). The goal of a logistic regression analysis is to find the best fitting, and most parsimonious probabilistic model to describe the relationship between an outcome and a set of predictors. What distinguishes the logistic regression model from the linear regression model is that the outcome variable in logistic regression is categorical. In fitting the logistic regression model, a maximum likelihood based approach is used to estimate the regression parameters (Hosmer and Lemeshow, 1989). Logistic regression analysis has been used for a QSAR modeling antibacterial study (Cronin et al., 2002) as well as in other areas. In the present study we have applied pharmacophore, recursive partitioning and logistic regression methods to aid in discriminating between MIC forming compounds and those which do not form an MIC.

A preliminary comparison between CYP3A4 ($+b_5$) MIC and non-MIC forming compounds based on the mean molecular weight showed a significant difference between the two groups of compounds

($p < 0.05$) indicating that larger molecules are more likely to form MIC in vitro with CYP3A4 ($+b_5$). Another comparison was done on compounds that were CYP3A4 substrates compared to non CYP3A4 substrates, which resulted in a similar outcome, i.e. there was a significant difference in the molecular weights between the two groups. This difference may be due to the specific set of data used in our study (the macrolide antibiotics and anti-HIV compounds tend to have higher molecular weights) and may not be a universal characteristic of substrates/inhibitors of CYP3A4. However there was no significant difference between the ClogP between the two groups of compounds. The qualitative approach of common feature pharmacophore alignments (Catalyst, 2003) for selected compounds suggests that CYP3A4 ($+b_5$) MIC formation requires at least multiple (four) hydrophobic interactions and a hydrogen bond acceptor interaction. Conversely, non-MIC forming compounds appear to possess fewer hydrophobic pharmacophore features with hydrogen bond donor rather than acceptor interactions (Figure 1). Molecules that are inactivators but apparently do not form an MIC also appear to be generally larger with more hydrophobic and hydrogen bond acceptor features that MIC and non-MIC forming compounds.

The recursive partitioning approach confirmed these findings to some extent, as a key descriptor in the ChemTree lowest rms scoring tree for the first split was C(N)-O(CC). This highlights a hydrophobic interaction while the second split C(CCC)-C(CN) is likely to represent a hydrogen bond acceptor interaction some distance from a hydrophobic feature. The key molecular descriptors used in all models as well as the summary of the training and test set correlations are described in Table 5. Similarly, in the single tree ChemTree model (Figure 3) selected for comparison with this multiple tree model, the descriptor C(CCC)-C(O) represents the hydrogen bond acceptor while the descriptor C(CCC)-N(CC) represents the hydrophobic interaction. When the different molecular descriptors generated in Cerius$^2$ CSAR were used with the ChemTree descriptors, different recursive partitioning models can be produced that suggest the radius of gyration is an important descriptor as well. This is

DMD Fast Forward. Published on May 30, 2007 as DOI: 10.1124/dmd.106.014613
This article has not been copyedited and formatted. The final version may differ from this version.



interesting as the initial Cerius$^2$ CSAR tree model does not include any splits based on this descriptor. The radius of gyration is calculated using the following equation:

$$\text{Radius of gyration} = \sqrt{\left( \Sigma \frac{\left(x_i^2 + y_i^2 + z_i^2\right)}{N} \right)}$$

where N is the number of atoms and x, y, z are the atomic coordinates relative to the center of mass (Cerius$^2$, 2003). When this descriptor is used, all twelve molecules with a radius of gyration less than 3.75 are not MIC forming, while a value between 3.75 and 3.95 suggests MIC formation for five compounds. This descriptor is however less discriminating for the 23 molecules with radius of gyration values above 3.95. Compared to the hydrogen bond and hydrophobic descriptors, radius of gyration is likely less important. While the ChemTree and Cerius$^2$ tree models classify the majority (10 of 12) of the literature test set molecules correctly it should also be noted that the addition of larger numbers of trees did not improve the prediction accuracy in this case. The logistic regression model only used a hydrogen bond acceptor descriptor and it possessed the highest prediction accuracy, (91.6 %), for the test set.

The linear and regression tree models for $k_{inact}$, $K_I$, and $k_{inact}/K_I$ prediction produced poor statistics for training and testing. In addition, there was no consistency on the selected molecular descriptors among the optimal prediction models. This potentially indicates either limitations in the feature set or the complexity of modeling such parameters with these methods alone.

The presence of an amine function does not always imply that a molecule will form a MIC (Kajita et al., 2002). For example, 3-hydroxy tamoxifen and 4-hydroxy tamoxifen do not form MIC while tamoxifen and n-desmethyltamoxifen form MIC in vitro (Zhao et al., 2002). Some macrolide antibiotics such as clarithromycin form MIC (Mayhew et al., 2000) while others like miocamycin do not (Kasahara et al., 2000), which suggests the amine may be sterically hindered in the later case such that it cannot orientate correctly in the CYP3A4 ($+b_5$) binding site. This type of interaction may not be picked up with the 2D-QSAR tree methods and may result in the misclassifications observed.

DMD Fast Forward. Published on May 30, 2007 as DOI: 10.1124/dmd.106.014613
This article has not been copyedited and formatted. The final version may differ from this version.



It would also appear that requirements for CYP3A4 ($+b_5$) mediated MIC formation requires other molecular properties besides primary, secondary, tertiary amines or methylenedioxyphenyl features. From this study, differentiating between MIC forming and non-MIC forming compounds is apparently complex, likely requiring numerous hydrophobic and hydrogen bond acceptor features for MIC formation. The pharmacophore for MIC forming compounds (Figure 1a) is similar to previously published pharmacophores for CYP3A4 inhibitors (Ekins et al., 1999a; Ekins et al., 2003b) and autoactivators (Ekins et al., 1999b) indicating that these binding sites may overlap to some extent. Compounds that inactivate but do not form MIC with CYP3A4 may possess all of these features but be too large overall. This would suggest some important size and molecular feature constraints on binding CYP3A4 ($+b_5$) which lopinavir, saquinavir, and nefazodone appear to exceed as they possess many more molecular features than MIC and non-MIC forming compounds. The non-MIC forming compounds indicate that there may be a minimal set of molecular features for MIC formation to occur. It is interesting to speculate that these features may impact the orientation of a molecule in the CYP3A4 binding site and the observation of MIC formation. For example the number of rotatable bonds correlates with the molecular weight ($r^2 = 0.68$, using the data for 54 molecules from Supplemental Table 1) and the number of hydrogen bond acceptors also correlates with the molecular weight ($r^2 = 0.75$, using the data for 54 molecules from Supplemental Table 1) and these larger compounds in turn tend to be MIC forming, upto a point. In contrast the correlation between molecular weight and the number of hydrogen bond donors is much lower ($r^2 = 0.42$ using the data for 54 molecules from Supplemental Table 1). This may indicate that molecules with greater conformational flexibility and size can position themselves optimally to form an MIC, while smaller more rigid molecules are less likely to be able to fulfill the key pharmacophore features. However, with increasing conformational flexibility of molecules could come the increased likelihood that fewer of the MIC formation events become observable with our current detection method, as many of the population of conformations do not result in MIC formation. A recent X-ray structure of the MIC forming compound

Erythromycin in CYP3A4 did not suggest the molecule was in a productive binding mode as the site of metabolism was too far from the heme (Ekroos and Sjogren, 2006), yet the authors did not indicate whether the orientation could result in the MIC.

It is important to note that this present study used molecular descriptors calculated on the parent molecule to predict MIC formation, when in fact a metabolite is actually binding to the heme. Our results would indicate that the descriptors selected from the parent molecules are transferable to capture the important features in the metabolites. This in itself may be valuable, as our approach therefore does not require the *a priori* knowledge or prediction of the metabolite/s involved. However once the metabolite/s is known it may be possible to use combined descriptors for the parent and metabolites as demonstrated previously for predicting N-dealkylation (Balakin et al., 2004). Future work may assess the chemical reactivity of the molecules that form MIC and attempt to determine the binding interactions with heme and elsewhere in the protein using site directed mutagenesis. In addition the models could be further tested with much larger diverse test sets, as those molecules selected in this study from the literature, represent six core structures.

An in vitro model was proposed that can predict the clinically observed inhibition of CYP3A4 after administration of clarithromycin, fluoxetine and diltiazem (Mayhew et al., 2000). In summary the in vitro rates for loss of CYP3A4 activity were used to predict the in vivo rates. Overall the in vitro-in vivo predictions were in accord with the clinical data. The clinical significance of being able to predict compounds that form MIC in humans is the constant polypharmacy that many humans subject themselves to on a daily basis. If one could predict *a priori* that a compound will form a MIC in humans and what the $k_{inact}$ value is, then it is possible to remove these types of structures from combinatorial libraries and screening collections. In this manuscript, models were developed and tested for prediction of compounds forming MIC with CYP3A4 ($+b_5$). Ultimately, based on the results in this study, the different QSAR models generated compliment the in vitro testing, representing a preliminary approach to predicting CYP3A4 ($+b_5$) MIC formation that may be improved with the

addition of further molecules to the training set, different descriptors or other algorithms. These approaches may be equally applicable to naturally occurring compounds as well as to drugs. The modeling methods could also be applied to other CYPs which are known to form MIC (Chatterjee and Franklin, 2003) such as CYP3A5 and CYP2D6, enabling a comparison with the CYP3A4 data presented here, providing insight into the features important for MIC formation and representing a more complete picture for pharmaceutical research.

DMD Fast Forward. Published on May 30, 2007 as DOI: 10.1124/dmd.106.014613
This article has not been copyedited and formatted. The final version may differ from this version.



## ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. Raymond Galinsky, who provided small amounts of MDMA, MDE, and MBDB for MIC assessment and Dr. Shikha Varma-O'Brien, (Accelrys Inc.) for assistance with Cerius$^2$ model descriptors.

**REFERENCES**

Balakin KV, Ekins S, Bugrim A, Ivanenkov YA, Korolev D, Nikolsky Y, Ivashchenko AA, Savchuk NP and Nikolskaya T (2004) Quantitative structure-metabolism relationship modeling of the metabolic N-dealkylation rates. *Drug Metab Dispos* **32**:1111-1120.

Breiman L, Friedman JH, Olshen RA and Stone CJ (1984) *Classification and regression trees*. Wadsworth.

Catalyst (2003) Version 4.8, Accelrys Inc., San Diego, CA.

Cerius$^2$ (2003) Version 4.8, Accelrys Inc., San Diego, CA.

Chatterjee P and Franklin MR (2003) Human cytochrome P450 inhibition and metabolic-intermediate complex formation by goldenseal extract and its methylenedioxyphenyl components. *Drug Metab Dispos* **31**:1391-1397.

Chen X, Rusinko I, A., Tropsha A and Young SS (1999) Automated pharmacophore identification for large chemical data sets. *J Chem Inf Comput Sci* **39**:887-896.

Chen X, Rusinko III A and Young SS (1998) Recursive partitioning analysis of a large structure-activity data set using three-dimensional descriptors. *J Chem Inf Comput Sci* **38**:1054-1062.

Crespi CL and Stresser DM (2000) Fluorometric screening for metabolism-based drug--drug interactions. *J Pharmacol Toxicol Methods* **44**(1):325-331.

Cronin MT, Aptula AO, Dearden JC, Duffy JC, Netzeva TI, Patel H, Rowe PH, Schultz TW, Worth AP, Voutzoulidis K and Schuurmann G (2002) Structure-based classification of antibacterial activity. *J Chem Inf Comput Sci* **42**(4):869-878.

de Graaf C, Vermeulen NP and Feenstra KA (2005) Cytochrome P450 in silico: an integrative modeling approach. *J Med Chem* **48**(8):2725-2755.

de Groot MJ, Ackland MJ, Horne VA, Alex AA and Jones BC (1999a) Novel approach to predicting P450-mediated drug metabolism: Development of a combined protein and pharmacophore model for CYP2D6. *J Med Chem* **42**:1515-1524.

de Groot MJ, Ackland MJ, Horne VA, Alex AA and Jones BC (1999b) A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed N-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *J Med Chem* **42**:4062-4070.

de Groot MJ and Ekins S (2002) Pharmacophore modeling of cytochromes P450. *Adv Drug Del Rev* **54**:367-383.

de Groot MJ, Vermeulen NPE, Kramer JD, van Acker FAA and Donné-Op den Kelder GM (1996) A three-dimensional protein model for human cytochrome P450 2D6 based on the crystal structures of P450 101, P450 102 and P450 108. *Chem Res Toxicol* **9**:1079-1091.

Ekins S, Berbaum J and Harrison RK (2003a) Generation and validation of rapid computational filters for CYP2D6 and CYP3A4. *Drug Metab Dispos* **31**:1077-1080.

Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH and Wrighton SA (1999a) Three and four dimensional-quantitative structure activity relationship analyses of CYP3A4 inhibitors. *J Pharm Exp Ther* **290**:429-438.

Ekins S, Bravi G, Wikel JH and Wrighton SA (1999b) Three dimensional quantitative structure activty relationship (3D-QSAR) analysis of CYP3A4 substrates. *J Pharmacol Exp Thera* **291**:424-433.

Ekins S, de Groot M and Jones JP (2001) Pharmacophore and three dimensional quantitative structure activity relationship methods for modeling cytochrome P450 active sites. *Drug Metab Dispos* **29**:936-944.

Ekins S, Ring BJ, Bravi G, Wikel JH and Wrighton SA (2000a) Predicting drug-drug interactions in silico using pharmacophores: a paradigm for the next millennium, in *Pharmacophore perception, development, and use in drug design* (Guner OF ed) pp 269-299, IUL, San Diego.

Ekins S, Ring BJ, Grace J, McRobie-Belle DJ and Wrighton SA (2000b) Present and future in vitro approaches for drug metabolism. *J Pharm Tox Methods* **44**:313-324.

Ekins S, Stresser DM and Williams JA (2003b) In vitro and pharmacophore insights into CYP3A enzymes. *Trends Pharmacol Sci* **24**:191-196.

Ekins S and Swaan PW (2004) Development of computational models for enzymes, transporters, channels and receptors relevant to ADME/TOX. *Rev Comp Chem* **20**:333-415.

Ekroos M and Sjogren T (2006) Structural basis for ligand promiscuity in cytochrome P450 3A4. *Proc Natl Acad Sci* **103**:13682-13687.

Ernest CS, 2nd, Hall SD and Jones DR (2005) Mechanism-based inactivation of CYP3A by HIV protease inhibitors. *J Pharmacol Exp Ther* **312**(2):583-591.

Franklin MR (1977) Inhibition of mixed-function oxidations by substrates forming reduced cytochrome P-450 metabolic-intermediate complexes. *Pharmacol Ther* **2**:227-245.

Gao F, Johnson DL, Ekins S, Janiszewski J, Kelly KG, Meyer RD and West M (2002) Optimizing higher throughput methods to assess drug-drug interactions for CYP1A2, CYP2C9, CYP2C19,

CYP2D6, rCYP2D6 and CYP3A4 *in vitro* using a single point IC50. *J Biomol Screen* **7**:373-382.

Hansch C and Zhang L (1993) Quantitative structure-activity relationships of cytochrome P-450. *Drug Metab Rev* **25**:1-48.

Hawkins DM, Young SS and Rusinko AI (1997) Analysis of large structure activity data set using recursive partitioning. *Quant Struct Act Rel* **16**:296-302.

Hosmer DW and Lemeshow S (1989) *Applied logistic regression*. Wiley, New York.

Jones JP, He M, Trager WF and Rettie AE (1996) Three-dimensional quantitative structure-activity relationship for inhibitors of cytochrome P4502C9. *Drug Metab Dispos* **24**:1-6.

Kajita J, Keiko I, Eiichi F, Kuwabara T and Kobayashi H (2002) Effects of olopatadine, a new antiallergic agent, on human liver microsomal cytochrome P450. *Drug Metab Dispos* **30**:1504-1511.

Kasahara M, Suzuki H and Izumi K (2000) Studies on the cytochrome P450 (CYP)-mediated metabolic properties of miocamycin: evaluation of the possibility of a metabolic intermediate complex formation with CYP, and identification of the human CYP isoforms. *Drug Metab Dispos* **28**:409-417.

Kim J-Y, Baek M, Lee S, Kim S-O, Dong M-S, Kim B-R and Kim D-H (2001) Characterization of the selectivity and mechanism of cytochrome P450 inhibition by dimethyl-4,4'-dimethoxy-5,6,5',6'-dimethylenedioxybiphenyl-2,2'dicarboxylate. *Drug Metab Dispos* **29**:1555-1560.

Lipinski CA, Lombardo F, Dominy BW and Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Del Rev* **23**:3-25.

Mayhew BS, Jones DR and Hall SD (2000) An in vitro model for predicting in vivo inhibition of cytochrome P450 3A4 by metabolic intermediate complex formation. *Drug Metab Dispos* **28**:1031-1037.

Pershing LK and Franklin MR (1982) Cytochrome P-450 metabolic-intermediate complex formation and induction by macrolide antibiotics; a new class of agents. *Xenobiotica* **12**(11):687-699.

Riley RJ, Parker AJ, Trigg S and Manners CN (2001) development of a generalized, quantitative physicochemical model of CYP3A4 inhibition for use in early drug discovery. *Pharm Res* **18**:652-655.

Silverman R (1988) Mechanism-based enzyme inactivation, in *Chemistry and Enzymology* pp 3-30, CRC Press, Boca Raton, FL.

Wang YH, Jones DR and Hall SD (2004) Prediction of cytochrome P450 3A inhibition by verapamil enantiomers and their metabolites. *Drug Metab Dispos* **32**:259-266.

Williams JA, Ring BJ, Cantrell VE, Jones DR, Eckstein J, Ruterbories K, Hamman MA, Hall SD and Wrighton SA (2002) Comparative metabolic capabilities of CYP3A4, CYP3A5, and CYP3A7. *Drug Metab Dispos* **30**:883-891.

Wrighton SA, Schuetz EG, Thummel KE, Shen DD, Korzekwa KR and Watkins PB (2000) The human CYP3A subfamily: practical considerations. *Drug Metab Revs* **32**:339-361.

Wu Y-J, Davis CD, Dworetzky S, Fitzpatrick WC, Harden D, He H, Knox RJ, Newton AE, Philip T, Polson C, Sivarao DV, Sun L-Q, Tertyshnikova S, Weaver D, Yeola S, Zoeckler M and Sinz MW (2003) Fluorine substitution can block CYP3A4 metabolism-dependent inhibition: Identification of (S)-N-[1-(4-fluoro-3-morpholin-4-ylphenyl)ethyl]-3-(4-fluorophenyl)acrylamide as an orally bioavailable KCNQ2 opener devoid of CYP3A4 metabolism-dependent inhibition. *J Med Chem* **46**:3778-3781.

Yamazaki H and Shimada T (1998) Comparative studies of in vitro inhibition of cytochrome P450 3A4-dependent testosterone 6b-hydroxylation by roxithromycin and its metabolites, troleandomycin, and erythromycin. *Drug Metab Dispos* **26**:1053-1057.

Young SS, Gombar VK, Emptage MR, Cariello NF and Lambert C (2002) Mixture deconvolution and analysis of Ames mutagenicity data. *Chemo Intell Lab Sys* **60**:5-11.

Zhao X-J, Jones DR, Wang Y-H, Grimm SW and Hall SD (2002) Reversible and irreversible inhibition of CYP3A enzymes by tamoxifen and metabolites. *Xenobiotica* **32**:863-878.

**FOOTNOTES**

a) Please send reprint requests to:

Stephen D. Hall, Ph.D.

Indiana University School of Medicine, Department of Medicine,

Division of Clinical Pharmacology, Wishard Memorial Hospital,

Myers Bldg W7123, Indianapolis, IN 46220,

E-mail: sdhall@iupui.edu

b) Supported by R01 AG13718, FDA FD-T-001756-01, and R01 GM74217

**FIGURE LEGENDS**

**Figure 1.** Metabolic intermediate complex formation of SKF-525A (A), phencyclidine (B), and lopinavir (C) with CYP3A4 (+b5). Lines represent the change in absorbance difference for scans at 5 (——), 15 (……), 30 (— — —) and 60 (—..—..—..) minutes.

**Figure 2**. Common feature pharmacophores for a. MIC forming compounds (SKF-525A, erythromycin, amprenavir, nor-verapamil), b Common feature pharmacophore for non-MIC forming compounds (mifepristone, sertraline, 4-hydroxy tamoxifen and paroxetine) and c. non-MIC forming compounds which inactivate CYP3A4 (lopinavir, saquinavir and nefazodone). Blue spheres = hydrophobic, green feature = hydrogen bond acceptor, Purple spheres = hydrogen bond donor.

**Figure 3**. Test set MIC forming compounds fitted to the MIC common features pharmacophore a. +beta-hydrastine, b. DDB, c. roxithromycin. Blue spheres = hydrophobic, green feature = hydrogen bond acceptor.

**Figure 4**. An example of a single automatically generated recursive partitioning tree for CYP3A4 ($+b_5$) MIC data is shown below. The parent population has 54 binary observations with a mean of 0.5 as half are MIC forming and half are non-MIC forming. The path length descriptor (Young et al., 2002) C(CCC)-C(O) is an integer distance range within the molecule and can be split significantly (P value = 7.77 e-7) two ways. From this point the remaining 40 compounds are split into 3 subsets using C(CCC)-N(CC).

**Figure 5.** Cerius 2 CSAR tree model for CYP3A4 ($+b_5$) MIC data. In this decision tree, a branch downwards indicates that the criteria were satisfied, that is, TRUE. A branch upwards indicates that the criteria were not satisfied at that particular decision point.

**Figure 6.** The R version 2.2.1 (www.r-project.org) tree model for CYP3A4 ($+b_5$) MIC formation.

**Table 1.** Estimated $k_{inact}$, $K_I$ and $k_{inact}/K_I$ generated with recombinant expressed CYP3A4 (+b5).

| Compound | $k_{inact}$ ($min^{-1}$) | $K_I$ (μM) | $k_{inact}/K_I$ ($min^{-1}$ $mM^{-1}$) | References |
|---|---|---|---|---|
| D-617 | 0.07 | 7.9 | 9 | (Wang et al., 2004) |
| N-desmethyltamoxifen | 0.13 | 12.0 | 11 | |
| Fluvoxamine | 0.05 | 3.7 | 13 | |
| Fluoxetine | 0.04 | 2.3 | 17 | (Mayhew et al., 2000) |
| Nortriptyline | 0.04 | 2.1 | 19 | |
| Nefazodone | 0.27 | 12.5 | 22 | |
| Clarithromycin | 0.23 | 4.1 | 56 | (Mayhew et al., 2000) |
| Tamoxifen | 0.03 | 0.5 | 59 | (Zhao et al., 2002) |
| R-verapamil | 0.39 | 6.5 | 60 | (Wang et al., 2004) |
| N-desmethylerythromycin | 0.34 | 5.7 | 60 | |
| Erythromycin | 0.17 | 2.3 | 75 | |
| 14-OH clarithromycin | 0.05 | 0.5 | 102 | |
| Amlodipine | 0.35 | 2.6 | 135 | |
| Nicardipine | 0.18 | 1.2 | 149 | |

| | | | | |
|---|---|---|---|---|
| Norverapamil | 1.12 | 5.9 | 190 | (Wang et al., 2004) |
| N-desmethyldiltiazem | 0.29 | 1.5 | 200 | (Mayhew et al., 2000) |
| S-verapamil | 0.64 | 3.0 | 215 | (Wang et al., 2004) |
| Lopinavir | 0.1 | 0.4 | 244 | (Ernest et al., 2005) |
| Propoxyphene | 0.41 | 1.4 | 293 | |
| Nelfinavir | 0.22 | 0.5 | 458 | (Ernest et al., 2005) |
| Diltiazem | 0.17 | 0.4 | 472 | (Jones et al., 1999) |
| SKF-525A | 0.25 | 0.4 | 625 | |
| triacetyloleandomycin | 0.11 | 0.1 | 1222 | |
| Mifepristone | 0.67 | 0.4 | 1634 | |
| Saquinavir | 0.31 | 0.2 | 1824 | (Ernest et al., 2005) |
| Amprenavir | 0.73 | 0.3 | 2808 | (Ernest et al., 2005) |
| Ritonavir | 0.32 | 0.1 | 3200 | (Ernest et al., 2005) |

All data were produced in this study except where explicitly referenced.

**Table 2.** Data on MIC formation with recombinant expressed CYP3A4(+b5) and used for training sets. CLogP and MWT were calculated with ChemDraw for Excel (CambridgeSoft, MA).

| Compound | Observed MIC [+] | 1 tree Predicted | 100 tree Predicted | CLogP | MWT | Reference |
|---|---|---|---|---|---|---|
| 14-OH-clarithromycin | 1 | 1 | 0.99 | 1.44 | 750 | |
| 3-OH-tamoxifen | 0 | 0.29 | 0.36 | 6.15 | 388 | (Zhao et al., 2002) |
| 4-OH-tamoxifen | 0 | 0.29 | 0.36 | 6.15 | 388 | (Zhao et al., 2002) |
| Amitriptyline | 0 | 0.29 | 0.28 | 4.85 | 277 | |
| Amphetamine | 0 | 0.29 | 0.15 | 1.74 | 135 | |
| Amprenavir | 1 | 1 | 1 | 3.6 | 520 | (Ernest et al., 2005) |
| Benzphetamine | 0 | 0.29 | 0.15 | 4.38 | 239 | |
| Brompheiramine | 0 | 0 | 0.02 | 2.91 | 335 | |
| Chlorpheniramine | 0 | 0 | 0.02 | 2.76 | 291 | |
| Clarithromycin | 1 | 1 | 0.99 | 2.37 | 748 | (Mayhew et al., 2000) |
| D-617 | 1 | 1 | 0.99 | 2.17 | 290 | (Wang et al., 2004) |
| Desacetyl-diltiazem | 1 | 1 | 0.99 | 2.96 | 372 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Desacetyl N-desmethyldiltiazem | 1 | 1 | 0.99 | 2.37 | 358 | |
| Desipramine | 1 | 1 | 0.8 | 4.47 | 266 | |
| Diltiazem | 1 | 1 | 0.99 | 3.65 | 415 | (Jones et al., 1999) |
| Diphenhydramine | 0 | 0.29 | 0.28 | 3.54 | 255 | |
| Erythromycin | 1 | 1 | 0.99 | 2.37 | 748 | |
| Fenfluramine | 0 | 0 | 0.15 | 3.3 | 231 | |
| Fluoxetine | 1 | 1 | 0.94 | 4.57 | 309 | (Mayhew et al., 2000) |
| Fluvoxamine | 1 | 1 | 0.76 | 3.32 | 318 | |
| Imipramine | 0 | 0.29 | 0.28 | 5.04 | 280 | |
| Indinavir | 1 | 0.29 | 0.36 | 3.68 | 614 | (Ernest et al., 2005) |
| Loperamide | 0 | 0.29 | 0.28 | 4.66 | 477 | |
| Lopinavir * | 0 | 0.29 | 0.36 | 7.05 | 657 | (Ernest et al., 2005) |
| MBDB | 0 | 0 | 0.02 | 2.38 | 207 | |
| MDE | 0 | 0 | 0.02 | 2.38 | 207 | |
| MDMA | 0 | 0 | 0.02 | 1.85 | 193 | |
| Meperidine | 0 | 0.29 | 0.15 | 2.23 | 247 | |
| Methamphetamine | 0 | 0 | 0.02 | 1.89 | 149 | |
| Methylphenidate | 0 | 0 | 0.24 | 2.56 | 233 | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Mifepristone | 0 | 0.29 | 0.28 | 4.46 | 430 | |
| Mirtazepine | 0 | 0.29 | 0.28 | 4.27 | 264 | |
| N-desmethylclarithromycin | 1 | 1 | 0.99 | 1.78 | 734 | |
| N-desmethyldiltiazem | 1 | 1 | 0.99 | 3.06 | 401 | (Mayhew et al., 2000) |
| N-desmethyltamoxifen | 1 | 1 | 0.71 | 6.22 | 358 | |
| Nefazodone * | 0 | 0.29 | 0.15 | 5.09 | 470 | |
| Nelfinavir | 1 | 1 | 0.8 | 5.84 | 568 | (Ernest et al., 2005) |
| Nicardipine | 1 | 1 | 0.96 | 2.2 | 481 | |
| Norfluoxetine | 1 | 0.29 | 0.64 | 4.3 | 295 | |
| Nortriptyline | 1 | 1 | 0.8 | 4.31 | 263 | |
| Nor-verapamil | 1 | 1 | 0.8 | 3.93 | 441 | (Wang et al., 2004) |
| Orphenadrine | 0 | 0.29 | 0.28 | 3.99 | 269 | |
| Paroxetine | 0 | 0 | 0.15 | 4.24 | 329 | |
| Phencylclidine | 0 | 0.29 | 0.15 | 5.1 | 243 | |
| Propoxyphene | 1 | 0.29 | 0.44 | 5.21 | 339 | |
| Ritonavir | 1 | 0.29 | 0.62 | 4.94 | 721 | (Ernest et al., 2005) |
| R-verapamil | 1 | 1 | 0.99 | 4.47 | 455 | (Wang et al., 2004) |
| Saquinavir * | 0 | 0.29 | 0.36 | 4.73 | 671 | (Ernest et al., 2005) |

| | | | | | | |
|---|---|---|---|---|---|---|
| Sertraline | 0 | 0 | 0.15 | 5.35 | 306 | |
| SKF-525A | 1 | 0.29 | 0.44 | 5.98 | 354 | |
| S-verapamil | 1 | 1 | 0.99 | 4.47 | 455 | (Wang et al., 2004) |
| Tamoxifen | 1 | 0.29 | 0.36 | 6.82 | 372 | (Zhao et al., 2002) |
| TAO | 1 | 0.29 | 0.48 | 5.25 | 798 | |
| Tranylcypromine | 0 | 0.29 | 0.28 | 1.25 | 133 | |

All data were produced in this study except where explicitly referenced. For predicted MIC formation, a value of > 0.5 was used for MIC formation, a value of < 0.5 for no MIC formation.

* molecules that do not form a MIC but inactivate CYP3A4

[+] 1 = molecule forms MIC, 0 = molecule does not form MIC

**Table 3.** Observed and predicted data for molecules in the validation test set.

| Name | Observed MIC | Chemtree 1 tree | Chemtree 100 trees | Cerius 2 tree | Chemtree and Cerius 2 descriptors | Logistic model trees | Logistic model regression |
|---|---|---|---|---|---|---|---|
| Hydrastinine | N | 0.29 | 0.28 | 0 | 0.2 | 0 | 0 |
| +beta-hydrastine | Y | 1 | 0.82 | 1 | 0.88 | 1 | 1 |
| - beta-hydrastine | Y | 1 | 0.82 | 1 | 0.88 | 1 | 1 |
| DDB+ | Y | 1 | 0.62 | 1 | 0.82 | 1 | 1 |
| Olopatadine | N | 0.29 | 0.29 | 0 | 0.43 | 0 | 0 |
| Olopatadine M1 | N | 1 | 0.8 | 0 | 0.61 | 0 | 0 |
| KCNQ2 opener * | N | 0 | 0.43 | 0 | 0.47 | 0 | 0 |
| Roxithromycin | Y | 1 | 0.99 | 1 | 0.99 | 1 | 1 |
| Roxithromycin M1 | Y | 1 | 0.99 | 1 | 0.94 | 1 | 1 |
| Roxithromycin M2 | Y | 1 | 0.99 | 0 | 0.99 | 1 | 1 |
| Roxithromycin M3 | Y | 1 | 0.99 | 0 | 0.99 | 1 | 1 |
| Miocamycin | N | 1 | 0.98 | 1 | 0.93 | 1 | 1 |

+ dimethyl-4,4'-dimethoxy-5,6,5',6'-dimethylenedioxybiphenyl-2,2'dicarboxylate

* (S)-N-[1-(4-fluoro-3-morpholin-4-ylphenyl)ethyl]-3-(4-fluorophenyl)acrylamide

**Table 4.** Coefficient of determination values for regression tree models and linear multiple regression used in the prediction of $k_{inact}$, $K_I$, and $k_{inact}/K_I$.

| Model | Set | $k_{inact}$ ($min^{-1}$) | $K_I$ ($\mu M$) | $k_{inact}/K_I$ ($min^{-1} \bullet nM^{-1}$) |
|---|---|---|---|---|
| Regression tree | Training | 0.41 | 0.39 | 0.64 |
| | Validation | 0.12 | 0.15 | 0.05 |
| Linear multiple regression | Training | 0.16 | 0.26 | 0.40 |
| | Validation | 0.23 | 0.26 | 0.11 |

**Table 5**. Summary of molecular descriptors selected and training and test set correlations for MIC predictions.

| Prediction Model | Selected Descriptors | Training ($r^2$) | Test set ($r^2$) |
|---|---|---|---|
| Pharmacophores | Hydrophobic feature and Hydrogen bond acceptor | NA | NA |
| ChemTree (1 and 100 tree models) | Hydrogen bond acceptor and hydrophobic | 49/54 = 91% | 10/12 = 83.3% |
| ChemTree 100 tree models with Cerius[2] descriptors | Hydrogen bond acceptor and hydrophobic and Radius of Gyration, | 49/54 = 91% | 10/12 = 83.3% |
| Cerius[2] CSAR Tree | Hydrogen bond acceptor, Hydrogen bond donor, area, and Dipole magnitude | 54/54 = 100% | 10/12 = 83.3% |
| Tree in R | Hydrogen bond acceptor, radius of gyration, sum of atomic polarization, and dipole magnitude | 52/54 = 96% | 11/12 = 91.6% |
| Linear model in R | Hydrogen bond acceptor, | 43/54 = 80% | 11/12 = 91.6% |

NA – not applicable

Figure 1

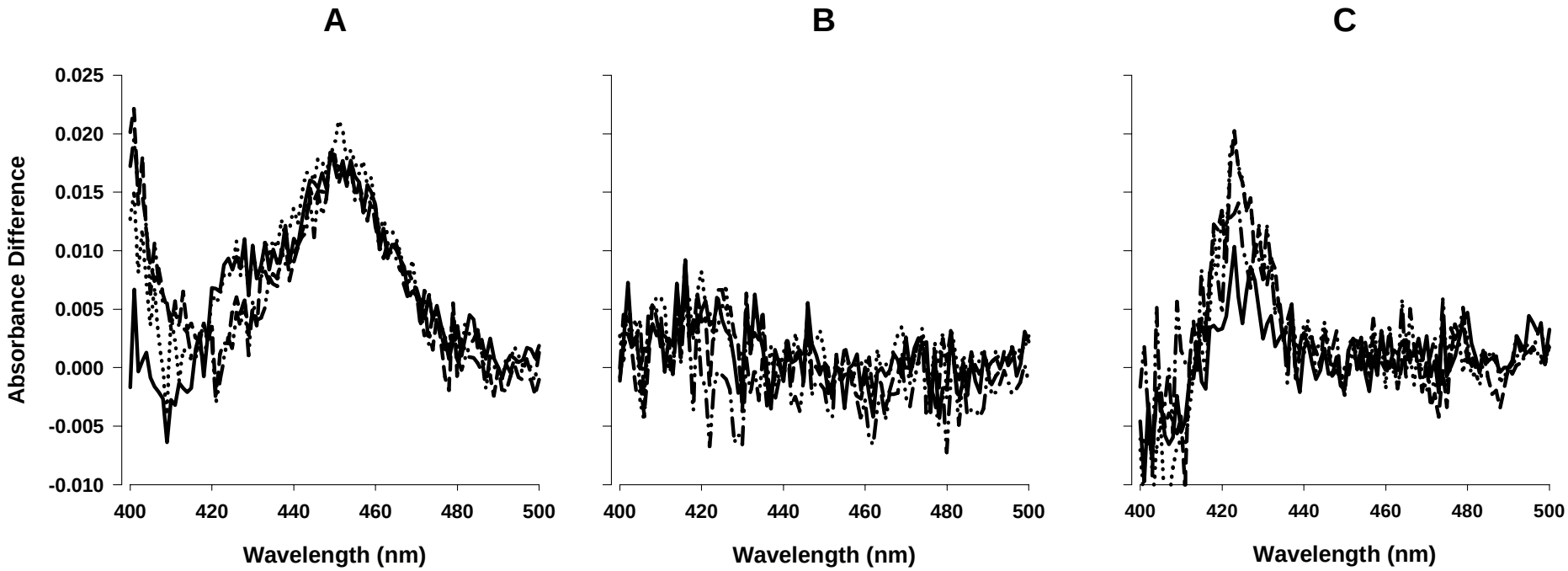
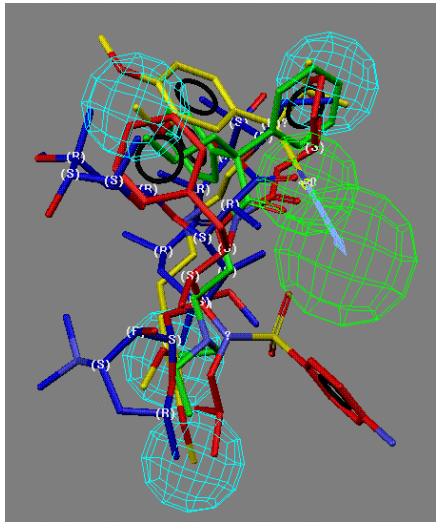
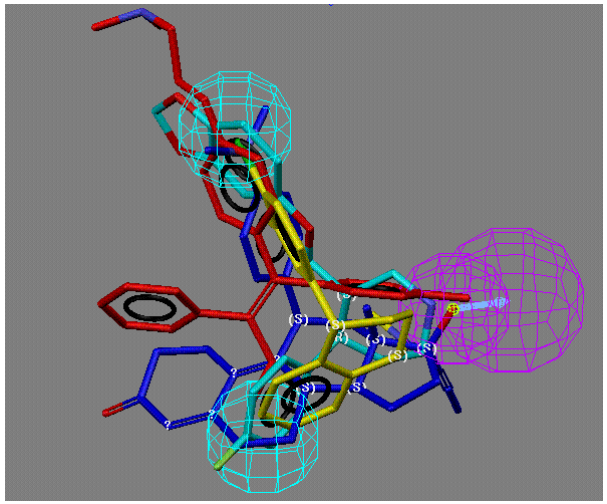


Figure 2

A.



B.



C.

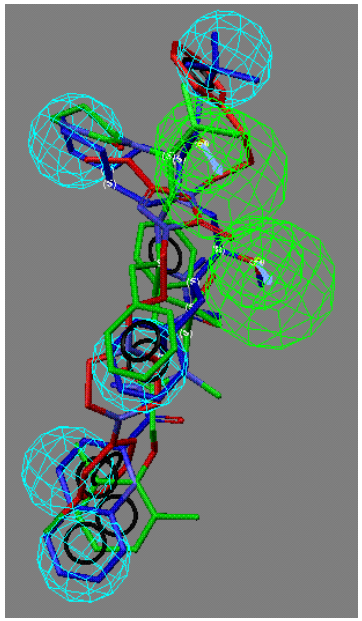
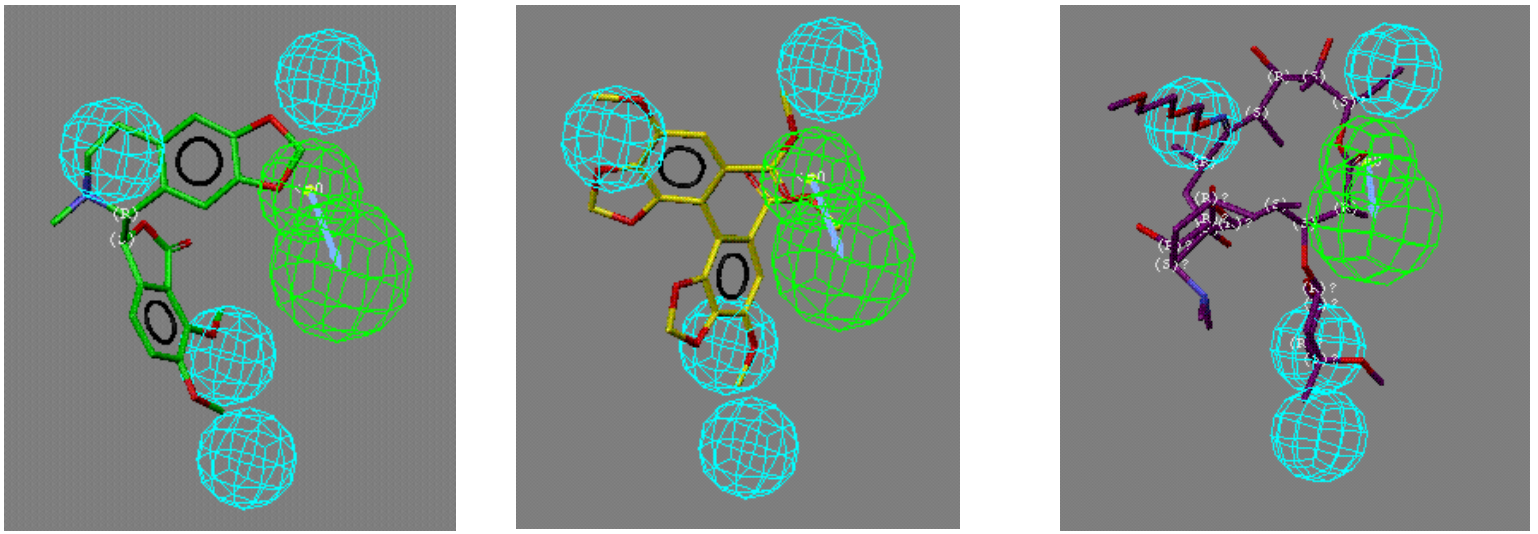


Figure 3



# Figure 4

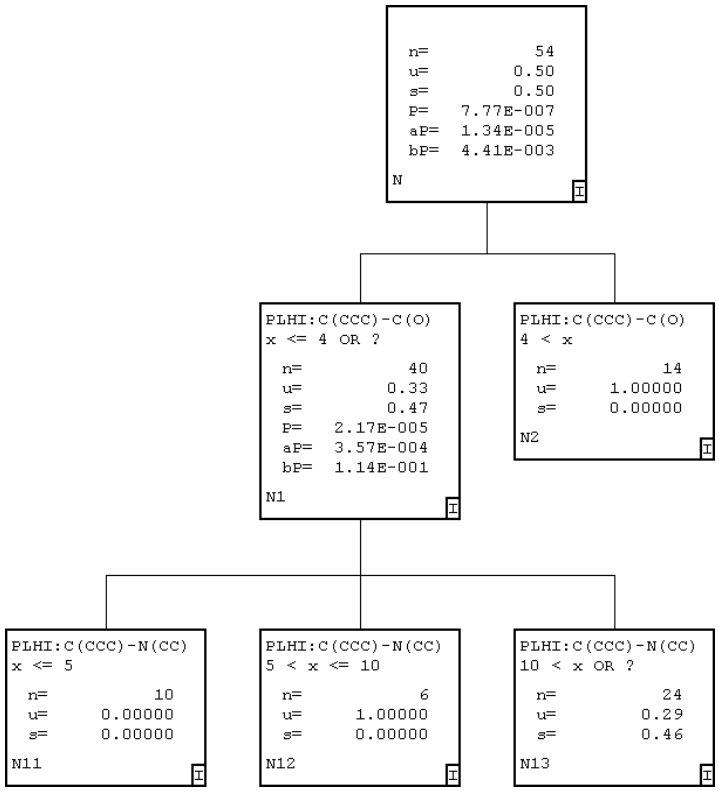


Figure 5

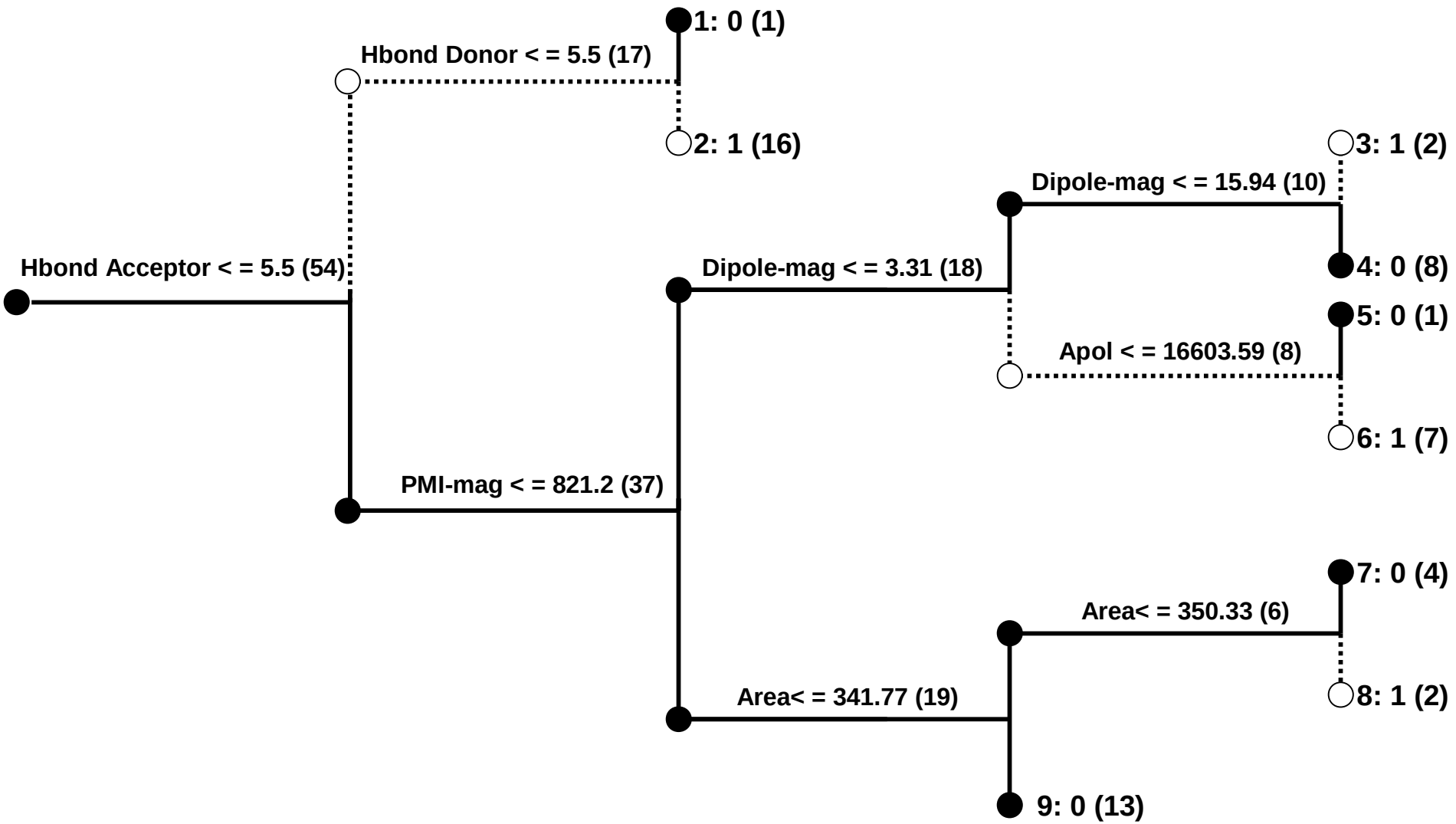


Figure 6