Bioorganic & Medicinal Chemistry 20 (2012) 2042–2053

Contents lists available at SciVerse ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Identification of cytochrome P450 2D6 and 2C9 substrates and inhibitors by QSAR analysis

Svava Ósk Jónsdóttir *, Tine Ringsted, Nikolai G. Nikolov, Marianne Dybdahl, Eva Bay Wedebye, Jay R. Niemelä

Department of Toxicology and Risk Assessment, National Food Institute, Technical University of Denmark, Mørkhøj Bygade 19, DK-2860 Søborg, Denmark

ARTICLE INFO

Article history:
Received 24 October 2011
Revised 21 January 2012
Accepted 25 January 2012
Available online 4 February 2012

Keywords:
QSAR
CYP2D6
CYP2C9
Substrate
Inhibitor
Procarcinogens

ABSTRACT

This paper presents four new QSAR models for CYP2C9 and CYP2D6 substrate recognition and inhibitor identification based on human clinical data. The models were used to screen a large data set of environmental chemicals for CYP activity, and to analyze the frequency of CYP activity among these compounds. A large fraction of these chemicals were found to be CYP active, and thus potentially capable of affecting human physiology. 20% of the compounds within applicability domain of the models were predicted to be CYP2C9 substrates, and 17% to be inhibitors. The corresponding numbers for CYP2D6 were 9% and 21%. Where the majority of CYP2C9 active compounds were predicted to be both a substrate and an inhibitor at the same time, the CYP2D6 active compounds were primarily predicted to be only inhibitors. It was demonstrated that the models could identify compound classes with a high occurrence of specific CYP activity. An overrepresentation was seen for poly-aromatic hydrocarbons (group of procarcinogens) among CYP2C9 active and mutagenic compounds compared to CYP2C9 inactive and mutagenic compounds. The mutagenicity was predicted with a QSAR model based on Ames in vitro test data.

1. Introduction

The cytochrome P450 (CYP) super family of heme containing enzymes plays part in the phase I metabolism of a wide range of endogenous compounds and xenobiotics,[1,2] including about 75% of marketed drugs.[3,4] The human genome encodes 57 different CYP genes, with five of these enzymes being responsible for the metabolism of 95% of drugs, namely CYP1A2, CYP2C9, CYP2C19, CYP2D6 and CYP3A4.[4–6] It is therefore an important factor in drug development and drug therapy to determine if a drug is metabolised by CYP enzymes.[3] In fact, improved understanding of CYP activity (metabolism, inhibition and induction) and the interplay between these processes is a key factor in identifying undesirable drug–drug interactions. The modulation of CYP activity by inhibition or induction of drugs or other chemicals can cause problems ranging from insufficient therapeutic effect to fatal toxic consequences.[7] Several drugs have been removed from the market due to inhibition of these enzymes by co-administrated drugs.[4–6]

Besides drugs, CYP enzymes detoxify environmental compounds and chemicals in consumer products. They also have the ability to form reactive intermediates which can damage DNA, lipids and proteins, and potentially lead to tumour initiation and cancer after long term exposure.[8]

The CYP2C9 and CYP2D6 isozymes are generally considered to be involved in the metabolism of 16–17% and 12–15% of prescribed drugs, respectively.[5,6,9] Both of these isozymes are primarily expressed in the liver and the small intestinal mucosa, with CYP2C9 being the CYP expressed in the largest abundance, after CYP3A4.[10,11] CYP2C9 is estimated to account for 15–20% of the total amount of CYPs expressed in the human liver on average, with the corresponding number for CYP2D6 being less than 2%.[3,12,13] Similarly, CYP2C9 and CYP2D6 are estimated to be 14% and 0.7% of the total amount of CYPs expressed in the human small intestinal mucosa on average,[10,11] and in smaller quantities in several other tissues.[14] CYP2D6 is also reported to be expressed in the brain.[15]

As discussed in detail in the recent literature, CYP2C9 and especially CYP2D6 are associated with substantial polymorphism causing large inter-individual differences in enzyme activity.[15] Of the Caucasians that have been genotyped for CYP2D6 activity, approximately 7% were identified as poor metabolizers, 10–15% as intermediate metabolizers and 5.5% as ultra-rapid metabolizers.[3,16,17] The poor metabolizer phenotype in the Asian population was found to be at about 1% and the ultra-rapid metabolizer phenotype

Abbreviations: CYP, cytochrome P450; CYP2C9, cytochrome P450 subtype 2C9; CYP2D6, cytochrome P450 subtype 2D6; CYP1A2, cytochrome P450 subtype 1A2; CYP2C19, cytochrome P450 subtype 2C19; CYP3A4, cytochrome P450 subtype 3A4; PLR, partial logistic regression; PLS, partial least-squares; EINECS, European Inventory of Existing Commercial Chemical Substances.

* Corresponding author. Tel.: +45 3588 6192; fax: +45 3588 7699.
E-mail address: sojo@food.dtu.dk (S.Ó. Jónsdóttir).

0968-0896/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2012.01.049

is reported to be present in over 20% in certain populations in North East Africa, Arabia and Oceania.[3]

It is clear that there exist rich possibilities of toxic consequences as a result of CYP2D6 metabolism or inhibition, due to the importance of CYP2D6 in drug metabolism, associated with relatively low expression in the liver, and the large inter-individual differences. This problem is also affected by the absence of known inducers of CYP2D6.[9] Thus pharmaceutical companies have developed strategies to reduce CYP2D6 inhibitory activity among lead candidates.[18] In addition to serious implications due to reduced, increased or alternative activity of prescribed drugs, the impact of other chemicals interacting with CYP2D6 is not fully understood.

Although CYP2C9 is expressed in greater abundance than CYP2D6, and its allelic variations have less profound influence on the enzyme activity, problems due to inter-individual differences of CYP2C9 occur as well. This can have fatal consequences for drugs like (S)-warfarin, that have a narrow therapeutic window and for which CYP2C9 constitutes a major metabolic pathway.[77] In recent years, genotyping with respect to CYP expression is increasingly used for providing personalized medicine for a variety of drugs, including (S)-warfarin.[19]

Polymorphic differences are estimated to cause moderately reduced CYP2C9 activity in ca. 14% of Caucasians and around 4% in Asians and Africans, and very low activity in 0.4% of Caucasians.[77] In addition to allelic variation, inter-individual differences can also be caused by specific nuclear receptors, like the constitutive androstane receptor (CAR) and the pregnane X receptor (PXR), which upon ligand binding up-regulate the expression of genes, including CYP2C9.

CYP2D6 is rather specific and mainly metabolizes basic molecules containing a protonatable nitrogen atom in 4–7 Å distance from the site of oxidation and coplanar aromatic rings. CYP2C9 on the other hand has a larger binding cavity and primarily metabolizes aromatic, lipophilic and either neutral or weakly acidic compounds that contain a hydrogen bond acceptor atom. CYP2C9 substrates are also characterized by a hydrogen bond donor site and/or an anionic moiety in 7–8 Å distance from the site of metabolism.[9,11,14,20,21] Some of the differences between substrates and inhibitors to these two CYPs have been illustrated by key pharmacophore features.[14,21,22] Freitas et al.[23] have identified key reference structures that discriminate CYP2C9 substrates from CYP2D6 substrates by the use of similarity searches, and provide 2D illustrations of typical substrates for these CYPs.

CYP2C9 metabolizes several endogenous compounds, such as arachidonic acid, linoleic acid, melatonin, retinoids and steroids,[24] as well as xenobiotics. This isozyme is also known to play a role in the metabolism of procarcinogens like for example, benzo[*a*]pyrene[25] which is found in coal tar, automobile exhaust fumes, tobacco smoke, wood smoke and in charcoal broiled food.[24] Another example is the procarcinogen dibenzo[*a*,*h*]antracene which has been shown to induce cancer in mice and rats.[26] Dibenzo[*a*,*h*]antracene is found in coal tar, automobile exhaust fumes, tobacco smoke, coke oven emissions, charcoal broiled food, vegetation near heavily travelled roads, surface water and soils near hazardous waste sites.[27]

Several QSAR studies for CYP-mediated metabolism have been published in the literature in recent years, particularly with focus on drug metabolism. These cover global QSAR models for both CYP inhibitor identification and substrate recognition, for example, models that discriminate CYP substrates from non-substrates and inhibitors from non-inhibitor, as well local QSAR models for predicting the biological activity for smaller and structurally similar data sets. A number of global and local models are discussed in recent reviews.[11,14,28–31]

Yap and Chen[32] assembled data sets based on clinical data and developed global consensus models for 3A4, 2D6 and 2C9 inhibitors and substrates using support vector machine (SVM). A new decision tree method was used to develop models on the same inhibitor data sets.[33] A partial least square (PLS)—regression tree (RT) consensus models for inhibition of the five most important CYPs for drug metabolism were based on 1500 compounds from Astra Zeneca's in-house database.[34] Hammann et al.[35] developed a set of models for 1A2, 2D6 and 3A4 substrates and inhibitors with several classification methods. Models for classifying substrates to five major CYPs based on considerable smaller data sets were published recently.[36] The above QSAR models were in general based on different types of physico-chemical, constitutional, topological, electrotopological, geometric and atomic charge based descriptors, among others.

Several global models for CYP2D6 substrate[37] and inhibitor[29,31,38–40] activity are available. A set of SVM and PLS models based on general molecular descriptors and structural fragments were developed for CYP2C9 inhibition.[41] QSAR models that discriminate between substrates selective to key CYPs,[42] as well as a model which aimed at identifying compounds that worked as inhibitors to several different CYP subtypes,[43] have also been proposed.

Several authors have used a combination of 3D modelling of protein–ligand interactions, 3D-QSAR and QSAR for predicting binding energies and to investigate the mechanism of inhibition for both CYP2C9[44,45] and CYP2D6.[46] Combined use of docking and QSAR analysis have shown to yield improved predictions of substrate binding energies for these two isozymes.[47,48]

Computational and knowledge based methods for identifying sites of metabolism (SOMs) and for predicting binding energies in CYP enzymes are discussed in recent reviews.[49–53] Docking methods have shown to provide useful information on SOMs, as well as on the mechanism of inhibition for both CYP2C9[54–56] and CYP2D6.[57–59]

QSAR models have been made for identifying the oxidation sites in 3A4, 2D6 and 2C9 substrates (regioselectivity). These models used the random forest method and descriptors that describe the local substructure and physico-chemical environment for each non-hydrogen atom in the molecule.[60] An interesting method, MetaSite, determines chemical fingerprints for the binding pockets of the CYP enzymes and corresponding fingerprints for potential substrates, and uses similarity analysis to identify substrate-CYP complimentarily. The MetaSite method, which is entirely based on the enzymes and the potential substrates 3D structures, has shown to be successful for correctly predicting SOMs and isoform selectivity.[61]

The aim of this study was to predict and analyse substrate recognition and inhibitor identification of the CYP2C9 and CYP2D6 isozymes among chemicals in our environment. For this purpose we developed a set of QSAR models, and screened a large inventory of chemicals with no available experimental data for CYP activity.

2. Material and methods

2.1. Data sets

The training sets for CYP substrates consisted of human clinical data for CYP2C9, and CYP2D6, primarily drugs, gathered from the literature,[32,60,62] and corresponding data for new drugs (http://dailymed.nlm.nih.gov/dailymed/about.cfm). The training sets for the corresponding CYP inhibitors consisted of human clinical data from the literature.[32] The data were either obtained from in vivo clinical experiments or various in vitro models like tissues slices, microsomes, cell cultures and purified and recombinant enzymes that have formed the basis for a clinical decision.[62] Epidemiological observations and case studies may also have served as input for such decisions. The negative data sets were composed of well-studied and extensively used drugs, which are known to be inhibitors/sub-

Table 1
The number of compounds in the respective full training sets and sub-sets for substrates/non-substrates and inhibitors/non-inhibitors

	CYP	Balance	Positives	Negatives	Total
All data					
Substrates	2C9	1:2.9	190	547	737
	2D6	1:2.0	246	500	746
Inhibitors	2C9	1:3.2	166	525	691
	2D6	1:2.8	180	511	691
Sub-sets					
Substrates	2C9 (set a)	1:1.4	190	273	463
	2C9 (set b)	1:1.4	190	274	464
	2D6 (set a)	1:1.4	246	355	601
	2D6 (set b)	1:1.4	246	355	601
Inhibitors	2C9 (set a)	1:1.6	166	262	428
	2C9 (set b)	1:1.6	166	263	429
	2D6 (set a)	1:1.4	180	255	435
	2D6 (set b)	1:1.4	180	256	436

The compounds were divided into sub-sets by random selection, and the sub-sets were used for training each of the sub-models.

strates/agonists to other CYPs, but with no reports given on the activity (substrate or inhibitor) for the relevant CYP.[32] This approach was used, because very few clinical reports that directly identify chemical substances as non-substrates or non-inhibitors exist.

The data were checked for structural replicates. In case of duplicate structures, one of the replicates was kept if the chemicals had the same activity and both were removed if they had different activity. The data were also searched for salts, and the structures were used in their non-ionized form in the training set. In addition, the structure of Bortezomib, which contains a boron atom, could not be imported into the Leadscope software and was therefore removed from the data set.

To obtain more equal balance with respect to the number of positive and negative compounds, all the training sets were divided into two sub-sets by random selection, see Table 1 for information on the overall data sets and sub-sets. In each case the two sub-sets were used separately to train the sub-models (models a and b) that were combined in an overall so-called ensemble model. Each sub-set was slightly imbalanced with 40% positives and 60% negatives. This was done to ensure a larger number of data points used for each sub-model and to obtain sufficiently high specificity to avoid unnecessary false positive predictions and at the same time to avoid low sensitivity. Thus our strategy was to develop models slightly higher in specificity than in sensitivity, and in such a way strengthen the quality of our positive predictions.

We experimented with dividing the data set into two sub-sets both by random selection and by hierarchical clustering based on structural features. However, it was seen that the overall performance of the models based on the randomly selected sub-sets was equally good.

2.2. Development of prediction models

The predictive QSAR models were developed with Leadscope Predictive Data Miner (Version 3.04-10, Leadscope, Inc., http://leadscope.com). All the molecular structures in the respective training sets were imported into Leadscope, and mined for the predefined structural features from Leadscopes template library by substructure analysis.[63] In each case, 10% or 13% of the structural features that best correlated with the response variable were selected with the manual feature selection option in Leadscope. The selection was done according to Yates' χ^2-test.[64] In addition, eight molecular descriptors, the octanol–water partition coefficient ($A\log P$), hydrogen bond acceptors, hydrogen bond donors, Lipinski score, atom count, parent compound molecular weight, polar surface area and rotatable bonds, were calculated for each structure. Redundant features were removed using the least redundant feature option in Leadscope. Predictive models were developed for each training set based on the identified set of the structural features and the molecular descriptors, using partial logistic regression (PLR). The number of optimal PLS (partial least-square) factors for each model was optimized using 10-fold (10 times 10% leave-out) cross-validation.

Each data set was re-modeled several times, where the percentage of the best features according to a χ^2-test was varied until the predictive performance measured by 10-fold cross-validation was not significantly improved. Different options for removing redundant features were also investigated.

For each overall training set, two separate sub-models were developed based on each of the sub-sets (see Table 1). The two sub-models were then combined into an overall ensemble model with the 'assemble average model' option in Leadscope, with both models assigned equal weights of 1.0.

The predictive performances of each sub-model, was evaluated by 10-fold cross-validation, and the specificity was also evaluated by external validation, using the portion of the negative data set not used for training the model.

To further evaluate the predictive performance and robustness of the models, 20% of the positive and negative data points were removed from the respective sub-sets by random selection. A new set of models was developed, with similar number of descriptors as the corresponding models developed for the full sub-sets. The structural features were selected based on the 80% of the data used for the model, the same eight molecular descriptors were added and redundant features were removed with the least redundant option. A new model was built in Leadscope in each case, and the model was used to predict the activity of the remaining 20% of the data set. It was verified in each case that the key features identified for the model based on 80% of the data were similar to the key features identified for the model built from the full sub-set. This was done in order to evaluate if a model with 80% of the data could be considered representative with respect to the corresponding model developed for the full sub-set.

2.3. Predictions within applicability domain

In order to make the predictions as reliable as possible, only predictions within the defined applicability domain were accepted. Class probabilities for positive predictions (p) were used for defining the domain, requiring $p \geqslant 0.7$ for positives and $p \leqslant 0.3$ for negatives, referred to as activity domain. In addition, it was required that a compound had at least 30% Tanimoto structural similarity with a training set compound to be considered in domain, referred to as structural domain. The Tanimoto similarity was calculated based on fingerprints of the Leadscope features used for each of the models. Compounds screened with the ensemble models were required to have 30% similarity with at least one training set compound in either sub-model a or b.

2.4. Screening a large inventory of chemicals for CYP activity

The resulting CYP models were used to screen a list of 51,067 chemicals from the EINECS list (European INventory of Existing Commercial Chemical Substances) in order to predict a percentage of potential CYP2C9 and CYP2D6 substrates and inhibitors present in environmental chemicals. The EINECS list contains discrete organic chemicals, inorganic compounds and other substances that were registered for use on the European market between January 1st, 1971 and September 18th, 1981.[65,66] Some of these chemicals are found in consumer products or used for other applications, and some of the compounds are not currently used in any products. As our modelling methods can only handle discrete organic

compounds, a sub-set of 51,067 molecules with a defined 2D structure, containing at least two carbon atoms and only certain atoms (H, Li, B, C, N, O, F, Na, Mg, Si, P, S, Cl, K, Ca, Br, I) was selected as our prediction set. The full EINECS list contains 100,204 compounds.

2.5. Substructure searches and correlations between CYP activity and a genotoxicity endpoint

The screening results were imported into our in-house database tool, built on Oasis DatabaseManager,[67] which contains model predictions for more than 150 adverse effects for up to 185,000 compounds, including the EINECS list of 51,067 compounds. The database contains search options to explore correlations between model predictions, as well as the possibility for substructure and structural similarity searches. Correlations among the different model predictions were explored, and the fractions of specific chemical groups within the data were investigated. Only predictions within domain of each of the predictive models were considered in this analysis.

Specifically, we have used our in-house database to search for correlations between predicted CYP actives and predicted Ames positives for specific classes of known procarcinogens. The database includes model predictions based on Ames in vitro mutagenicity data. The Ames model was developed with the MultiCASE software based on a training set of 4102 chemicals, 2299 positive and 1803 negative. 5 times twofold 50% cross-validation gave sensitivity of 84%, specificity of 83% and concordance of 84% within domain of the model, which is close to the observed experimental variability for the Ames in vitro test.[68]

The substructure search facility incorporated into the system, was used to investigate the importance of CYP2C9 and CYP2D6 interactions with respect to chemical classes known to contain chemical procarcinogens, including poly-aromatic (polycyclic aromatic) hydrocarbons, aromatic amines, heterocyclic amines and nitrosamines.[69]

3. Results and discussions

3.1. Model training and validation

The training and validation performances of the developed QSAR models are shown in Table 2. The four QSAR models presented discriminate CYP2C9 and CYP2D6 substrates from non-substrates and CYP2C9 and CYP2D6 inhibitors from non-inhibitors, respectively. Each of these models was comprised of two different sub-models that were combined to an overall ensemble model.

The cross-validation and external validation results indicate that the models generally have good predictive performances with overall concordances between 79% and 83%, specificities from 80% to 88% and sensitivities between 66% and 73%. As the sub-sets used for training the models were slightly imbalanced towards a larger portion of negatives in each set, all the models have somewhat higher specificities compared to sensitivities. This ensures a lower probability of making false positive predictions compared to false negative predictions, an approach used to avoid overestimating CYP activity.

As the data sets are fairly small and relatively structurally diverse, we have evaluated the predictive performance of the QSAR models with a combination of 10-fold cross-validation of the full data set models, and external validations for the part of the negative compounds not used for developing the given sub-model. This was done in order to develop models that covered as large a structural domain as possible. The variability in the cross-validated concordances of the individual sub-models in Table 2 were estimated by repeating the 10-fold cross-validation 5 times by Leadscope, with different random division into subsets by Leadscope. The same set of descriptors was used in each case, but the number of PLS factors could vary. The variability in concordance was found generally to be in the range 0.5–1.5%.

Another set of models was used to better investigate the predictivity and the robustness of the data. These models had 20% of both the positives and the negatives of each sub-set removed for external validation (Table 3). The mean external validation performances for the two sub-models with structural and activity domain applied were similar to the corresponding mean 10-fold cross-validation performances. Taking into account the statistical uncertainties imposed by selecting sub-sets of compounds from a relative small and structurally diverse data set at random, both the external validation results within domain and the cross-validation performances showed that robust prediction models were developed. It was also seen from the external validation results that the predictive performances were significantly improved when structural and activity domain definitions were used. These results show that it is important to apply a domain in order to secure appropriate predictive performance. Thus only predictions with at least 30% structurally similarity to at least one training

Table 2
Model performances of each of the Leadscope models developed for CYP2C9 and CYP2D6 substrates and inhibitors

	Best features (%)	No. Descr./PLS factors	Training performance			10-fold cross-validation			External validation	
			Sensitivity (%)	Specificity (%)	Concordance (%)	Sensitivity (%)	Specificity (%)	Concordance (%)	Specificity no domain (%)	Specificity in domain (%)
Substrates										
CYP2C9 ensemble			77.9	86.4	84.2	70.0	82.2	79.1		
- model a	10	167/3	71.1	90.1	82.3	66.3	83.8	76.6	75.5	75.8
- model b	10	163/2	66.3	82.8	76.1	62.1	80.7	73.1	73.6	78.7
CYP2D6 ensemble			78.0	87.0	84.0	73.2	84.4	80.7		
- model a	13	222/2	77.2	84.5	81.5	72.8	83.3	79.0	72.4	81.5
- model b	13	238/2	74.8	87.6	82.3	69.9	84.7	78.7	84.1	88.5
Inhibitors										
CYP2C9 ensemble			72.3	90.7	86.3	66.3	88.4	83.1		
- model a	10	152/2	70.5	90.5	82.7	67.5	88.2	80.1	81.4	83.1
- model b	10	165/2	70.5	90.9	83.0	66.9	88.6	80.2	81.7	86.0
CYP2D6 ensemble			76.1	88.3	85.1	68.9	84.5	80.5		
- model a	10	153/2	73.9	88.6	82.5	65.6	85.9	77.5	82.4	81.2
- model b	10	169/2	73.9	87.9	82.1	67.8	83.2	76.8	74.5	81.1

In each case two different sub-models were developed, with the two sub-models based on all the positive compounds in the data set, and two different portions of the negative compounds. Percent best structural features extracted from the Leadscope library, the number of descriptors (structural features and molecular descriptors) and PLS factors of each model are listed. The predictive performances were evaluated by 10-fold cross-validation and external validation of the portion of the negative set not used to develop a given sub-model.

Table 3
Cross-validation performances of the Leadscope models developed for 80% of the training sets for CYP2C9 and CYP2D6 substrates and inhibitors, and external validation results for the remaining 20% of the data set in each case

	10-fold cross-validation			External validation no domain			External validation in domain		
	Sensitivity (%)	Specificity (%)	Concordance (%)	Sensitivity (%)	Specificity (%)	Concordance (%)	Sensitivity (%)	Specificity (%)	Concordance (%)
Substrates									
CYP2C9: mean	65.1	80.3	74.1	60.5	75.5	69.4	67.8	86.1	79.0
CYP2D6: mean	71.9	83.6	78.8	57.1	78.9	70.0	67.4	83.4	77.2
Inhibitors									
CYP2C9: mean	67.0	89.3	80.6	50.0	79.2	68.0	71.8	88.3	83.1
CYP2D6: mean	71.2	85.4	78.7	56.9	76.5	68.4	63.7	79.2	72.3

. External validation performances are shown with and without the use of model domain. The mean values for the two different sub-models are listed

Table 4
The best correlated positive and negative features for the CYP2C9 substrate sub-sets, evaluated by χ^2 statistics. The number of positive (Np) and negative (Nn) compounds that contain each feature are listed

Key features for 2C9 substrates, set a	χ^2	Nn	Np	Key features for 2C9 substrates, set b	χ^2	Nn	Np
Carboxylic acid, alkyl, acyc–	11.2	8	21	Methane, 1-aryl-,1-carbonyl–	11.5	2	13
Benzopyran	8.7	2	11	Pyridine, 2-amino–	9.4	0	8
Carboxylic acid, alkyl–	7.4	12	22	Methane, 1-carbonyl-,1-phenyl–	8.8	2	11
Ether, phenyl–	6.9	42	49	Sulfonyl group	8.4	16	27
Alcohol, alkenyl–	6.8	3	11	Alcohol, alkenyl–	6.9	3	11
Sulfonamide, *N*-alkyl–	6.8	1	8	Benzopyran	6.9	3	11
Indole	6.8	8	17	1-Benzene-sulfonamide, 4-alkyl–	6.5	0	6
Sulfonamide	6.4	14	23	Sulfonamide	5.7	15	23
Methane, 1-alkenyl-,1-phenyl–	6.4	0	6	Benzene, 1-sulfonyl–	5.7	11	19
Benzene, 1-methyl–	5.6	10	18	Benzene, 1-alkenyl-,4-alkoxy–	5.5	1	7
Benzene, 1-sulfonyl–	5.6	11	19	Pyridine, 2-alkylamino–	5.0	0	5
Ether, ethyl–	5.4	1	7	Carboxylic acid, alkyl, acyc–	4.9	14	21
Benzopyran, 4-oxo–	5.4	1	7	1-Benzene-sulfonamide	4.0	10	16
1-Methane-carboxylic acid, 1-phenyl–	5.4	1	7	Benzopyran, 4-oxo–	3.7	2	7
Methane, 1-carbonyl-,1-phenyl–	5.3	4	11	1-Methane-carboxylic acid, 1-phenyl–	3.7	2	7
Benzene, 1-oxy–	4.9	61	61	Benzopyran, 4-oxo, 2-hydroxy–	3.6	0	4
Piperazine, 1-(alkyl, acyc)–	4.9	7	14	Indole, 1-(alkyl, acyc)–	3.6	0	4
Pyrrole	4.7	10	17	Indole, 3-carbonyl–	3.6	0	4
Sulfonyl group	4.4	21	27	Pyrazolidine, 1-phenyl–	3.6	0	4
Pyridine(H)	20.4	65	14	Pyridine(H)	29.3	77	14
Pyridine(H), 3-carbonyl–	14.4	22	0	Piperidine	15.7	50	1
Pyridine(H), 2-(alkyl, acyc)–	12.1	19	0	Pyridine(H), 3-carbonyl–	11.3	18	0
Oxycarbonyl, *O*-alkyl–	11.7	50	13	Carboxylate, alkyl, cyc–	10.5	17	0
Carboxylate, alkyl, cyc–	11.4	18	0	Pyridine(H), 1-(alkyl, acyc)–	9.9	39	9
Pyridine, 1,4-dihydro, 3-carbonyl–	11.4	18	0	*tert*-Amine	9.4	108	48
Carboxylate	10.5	44	11	Amine, alkenyl, cyc–	9.1	21	2
Oxycarbonyl, *O*-(alkyl, acyc)–	10.4	28	4	Carbonyl, hydroxymethyl–	9.0	18	1
Amine, alkenyl, cyc–	9.2	21	2	Carboxylate, alkyl–	8.7	28	5
Nitro	8.7	23	3	Alkene, trisubst	8.6	39	1
Carboxylate, alkyl–	8.1	27	5	Carboxylate	8.6	41	11
Piperidine	7.6	37	1	Carbonyl, oxymethyl–	8.3	32	7
Pyridine, 1,4-dihydro, 4-phenyl–	7.0	12	0	Oxycarbonyl, *O*-alkyl–	8.0	44	13
Oxycarbonyl, *O*-methyl–	6.9	15	1	Steroid, 3-oxo–	7.7	19	2
tert-Amine	6.3	100	48	Diene, 1,3–	7.5	16	1
Pyran(H), 4-oxy–	6.2	11	0	*sec*-Amine(NH)	7.1	50	17
sec-Amine(NH)	5.8	47	17	Oxycarbonyl, *O*-(alkyl, acyc)–	7.0	23	4
Alcohol, cyclohexyl–	5.7	16	2	Pyridine(H), 2-(alkyl, acyc)–	6.2	11	0

set compound, as well as predictions with 70% or higher probability of being correct in these Leadscope models were accepted. The overlap was high between the structural features selected in the models based on 80% of the data and in the corresponding models for the full data sets.

The variability's in the mean concordances given in Table 3 were evaluated by calculating the standard deviations of the concordances of each of the two underlying sub-models. Although standard deviations for two numbers are associated with some uncertainty, this analysis gives an indication of the variability between two sub-models, developed for two different sub-sets of the data. The standard deviations of the 10-fold cross-validated concordances from Leadscope were between 1.2% and 2.6% for the four models. The standard deviations in the concordances obtained from external validation of two 20% leave-out sets were around 5% for three of the models, and 14% for the CYP2D6 inhibitor model. The variability is both influenced by the random selection of the leave-out sets and the diversity of the two sub-sets.

The developed QSAR models have generally performances similar to the previously published global models based on data sets of similar size. Yap and Chen[32] have obtained considerably better performances in their models using a combination of up to 80 SVM models. We have, however chosen to use a simpler and more easily interpretable modeling scheme with respect to relating CYP activity to certain structural features, using two sub-models and a linear PLR method. Thus a battery of models of good accuracy was obtained, based on structural features relevant to specific characteristic of CYP2D6 and CYP2C9 (discussed in the following section).

Table 5
The best correlated positive and negative features for the CYP2C9 inhibitor sub-sets, evaluated by χ^2 statistics

Key features for 2C9 inhibitors, set a	χ^2	Nn	Np	Key features for 2C9 inhibitors, set b	χ^2	Nn	Np
Sulfonamide, *N*-aryl–	20.3	0	14	1-Benzene-sulfonamide, 4-amino–	18.7	0	13
1-Benzene-sulfonamide	16.4	7	22	1-Benzene-sulfonamide	18.3	6	22
1-Benzene-sulfonamide, 4-amino–	15.6	1	13	Benzene, 1-amino(NH_2)–	13.1	2	13
Benzene, 1-sulfonyl–	13.0	10	23	Sulfonamide	9.5	16	26
thiophene	12.3	1	11	Benzene, 1-alkyl-,2-halo–	9.4	3	12
Benzene, 1-allyl–	10.0	5	15	Sulfoxide, alkyl–	8.8	0	7
benzopyran	7.7	1	8	Sulfonamide, *N*-aryl–	7.3	6	14
Pyridine, 3-(alkyl, acyc)–	7.7	1	8	Imidazole	7.3	13	21
Benzene, 1-alkyl-,2-halo–	6.2	5	12	Benzopyran, 4-oxo–	7.2	0	6
Methane, 1-carbonyl-,1-carbonylamino–	5.7	2	8	Sulfonyl group	7.0	20	27
Isoxazole, 5-(alkyl, acyc)–	5.6	0	5	Benzene, 1-alkyl-,4-aryl–	6.2	1	7
carboxylate, alkenyl–	5.0	6	12	Benzopyran	5.7	2	8
Benzene, 1-amino(NH_2)–	5.0	7	13	Methane, 1-carbonyl-,1-carbonylamino–	5.7	2	8
Alkene, tetrasubst	4.7	1	6	Benzene, 1-alkoxy-,4-(3-oxopropyl)–	5.6	0	5
Benzopyran, 4-oxo–	4.7	1	6	sulfoxide, aryl–	5.6	0	5
Ethane, 1-carbonylamino-,2-phenyl–	4.7	1	6	Benzene, 1-alkyl-,4-fluoro–	4.8	1	6
				1-Methane-carboxylic acid, 1-phenyl–	4.8	1	6
tert-Amine	19.1	114	37	Carboxylic acid, alkyl, acyc–	13.7	7	2
Pyridine(H), 2-(alkyl, cyc)–	10.7	22	1	*sec*-Amine(NH), alkyl–	13.4	43	7
Carbonyl, alkyl, cyc–	10.1	70	22	*tert*-Amine	11.9	102	37
Benzene, 1-alkylamino–	9.8	32	5	Ketone, alkyl, cyc–	10.1	30	4
Carboxylic acid, alkyl, acyc–	9.3	10	2	Benzene, 1-alkylamino–	9.7	32	5
Ether, alkyl, cyc–	9.3	41	9	Carbonyl, alkyl, cyc–	9.5	69	22
Ketone, alkyl, cyc–	8.9	28	4	Steroid	9.2	23	2
tert-Amine, aryl–	8.6	35	7	Steroid, 17-hydroxy–	8.2	15	0
Carbonyl, oxymethyl–	8.3	27	4	Piperidine	7.4	42	11
Pyrrolidine	8.1	24	3	Carboxylic acid, alkyl–	7.2	12	2
Piperidine	7.4	42	11	Benzene, 1-alkyl-,2-chloro-	6.7	3	1
Pyridine(H)	7.3	66	23	Alcohol, cyclohexyl–	6.7	16	1
Pyridine(H), 1-(alkyl, acyc)–	6.7	34	8	Carbonyl, oxymethyl–	6.5	24	4
Carbonyl, hydroxymethyl–	6.7	16	1	Benzene, 1-(alkyl, acyc)–	6.4	74	67
Alcohol, alkenyl–	6.6	3	1	Steroid, 3-oxo–	6.1	18	2
Pyridine(H), 4-oxo–	6.2	12	0	Pyridine(H), 2-(alkyl, cyc)–	6.0	15	1
Carboxylic acid, alkyl–	6.2	13	2	Pyridine(H)	5.9	63	23
Steroid	6.1	18	2	Pyridine(H), 1-(alkyl, acyc)–	5.7	32	8
Amine, allyl–	6.1	15	1	Glycine	5.6	11	0
sec-Amine(NH), alkyl–	5.8	30	7	Carboxamide (NR2)	5.5	36	1

The number of positive (Np) and negative (Nn) compounds that contain each feature are listed.

3.2. Analysis of structural features and molecular descriptors

Tables 4–7 show the best correlating positive and negative features evaluated by χ^2 statistics. It is seen by looking at the occurrences of each feature in the positive and negative data sets, that the models were based on a combination of many important features, and not dominated by a few particularly descriptive structural features. This is not surprising as the CYP enzymes bind a variety of different molecular structures. A number of sulfonamide and sulfonyl containing features (Tables 4 and 5) were prominent among the positive features for both the CYP2C9 substrates and inhibitors, reflecting the CYP2C9 receptors selectivity towards such compounds. Features containing carboxylic acid and carbonyl groups were also primarily seen among the positive features for CYP2C9 substrates, and mostly in the list of negative features for CYP2C9 inhibitors. Likewise various pyridine and piperdine containing features were found among the corresponding negative features for both substrates and inhibitors, with pyridine(H) and piperidine being particularly good discriminators for the non-substrates. Several carboxylate and oxycarbonyl structures were prominent for the non-substrates, but not for the non-inhibitors.

The CYP2D6 substrates and inhibitors (Tables 6 and 7) were characterized by positive features for the tertiary and secondary amine groups, and pyridine, piperidine and benzene rings, with especially many positive features for inhibitors containing a benzene ring. Several non-aromatic positive features were also found for the inhibitors. The morphinan feature appeared on the positive list for substrates, which could be explained by the many opiods metabolized by CYP2D6. The identified features illustrate the importance of aromatic amines and other nitrogen containing compounds among the CYP2D6 substrates. However, the number of benzene rings within the positive features for CYP2D6 inhibitors is somewhat surprising, as the inhibitor data was seen to contain a significantly smaller portion of aromatic hydrocarbons than the substrate data. Various functional groups containing carbonyl dominated the negative features for both CYP2D6 substrates and inhibitors, as well as carboxamide, steroid and sulfonamide features. Carboxylic acid and oxycarbonyl groups were also seen within the key negative feature list for the substrates, but not for the inhibitors.

Thus it can be seen that the key features used in the QSAR models indeed reflect the selectivity of CYP2C9 towards slightly acidic compounds and the selectivity of CYP2D6 towards basic molecules that contain protonatable nitrogen atoms.[9,20] The best reference structures that discriminate between CYP2C9 and CYP2D6 substrates, identified by Freitas et al.[23] also illustrate the same tendencies by 2D representations. Thus it can be concluded that the developed QSAR models are based on structural features that are relevant with respect to the respective CYP activity.

The differences between the substrate and inhibitor data sets were also examined by calculating the mean and standard deviation for six of the molecular descriptors, included in the descriptor set (Fig. 1). For both CYP2C9 and CYP2D6 the inhibitors were seen to have larger mean molecular weight, larger number of hydrogen bond acceptor atoms, larger number of rotatable bond and larger polar surface area than the substrates. The mean number of hydrogen bond donor atoms and octanol–water partition coefficients were also slightly larger for the inhibitors. The molecular

Table 6
The best correlated positive and negative features for the CYP2D6 substrate sub-sets, evaluated by χ^2 statistics

Key features for 2D6 substrates, set a	χ^2	Nn	Np	Key features for 2D6 substrates, set b	χ^2	Nn	Np
tert-Amine	28.5	100	123	tert-Amine, alkyl-	31.8	96	123
Piperidine, 4-aryl–	27.2	1	22	Piperidine, 4-aryl-	27.2	1	22
Piperidine, 1-(alkyl, acyc)–	21.5	14	37	*sec*-Amine(NH), alkyl-	19.7	29	52
Piperidine, 4-(alkyl, cyc)–	20.0	2	19	Piperidine, 4-phenyl-	19.7	1	17
Piperidine	19.7	30	53	Piperidine, 1-(alkyl, acyc)-	18.7	16	37
Piperidine, 4-phenyl–	19.7	1	17	Benzene, 1-alkoxy-,4-alkyl-	16.7	14	33
Benzene, 1-alkyl-,2-(2-aminoethyl)–	18.5	2	18	Benzene, 1-alkyl-,2-(2-aminoethyl)-	16.1	3	18
Azepine	18.2	0	14	Ethane, 1-alkylamino-,2-hydroxy-	16.0	8	25
sec-Amine(NH), alkyl–	16.7	32	52	Piperidine, 4-(alkyl, cyc)–	15.4	4	19
Ethane, 1-alkylamino-,2-hydroxy–	16.0	8	25	Benzene, 1-oxy–	14.1	80	91
Benzene, 1-alkoxy-,4-alkyl–	15.4	15	33	Morphinan	13.7	0	11
Benzene, 1,2-dialkyl–	15.0	11	28	Amine(NH), methyl, alkyl–	12.8	2	14
Piperidine, 2-(alkyl, cyc)–	14.0	4	18	Sulfide, phenyl–	12.8	2	14
Morphinan	13.7	0	11	Azepine	12.8	2	14
Benzene, 1-(2-aminoethyl)–	13.7	29	46	Piperidine	12.4	38	53
Benzene, 1-oxy–	13.5	81	91	Benzene, 1-alkoxy-,4-(2-aminoethyl)–	12.2	5	18
Amine(NH), methyl, alkyl–	12.8	2	14	Pyridine(H), 1-(alkyl, acyc)–	11.8	26	41
Pyridine(H), 4-aryl–	11.6	11	25	Benzene, 1,2-dialkyl–	11.2	14	28
Carbonyl, alkenyl, cyc–	30.1	73	11	Carbonyl, alkenyl, cyc–	29.3	72	11
Carboxy	27.6	49	3	Carboxy	23.7	44	3
Ketone, alkenyl–	24.6	48	4	Oxycarbonyl, *O*-alkyl–	19.1	69	16
Carbonyl, alkyl, acyc–	16.7	91	29	Ketone, alkenyl–	17.1	38	4
Carbonyl, alkyl, cyc–	16.3	94	31	Oxycarbonyl, *O*-(alkyl, cyc)–	15.9	44	7
Oxycarbonyl, *O*-alkyl–	15.8	64	16	Carboxylate	15.9	62	15
Ketone	14.1	77	24	Carbonyl, alkyl, cyc–	15.7	93	31
Carboxamide	13.6	100	37	Carboxylic acid, alkyl, acyc–	15.4	27	1
Oxycarbonyl, *O*-(alkyl, cyc)–	13.2	40	7	Carboxylate, alkyl–	13.6	43	8
Carboxylic acid, alkyl, acyc–	13.2	24	1	Amine, alkenyl, cyc–	12.9	32	4
Carbonyl, oxymethyl–	12.8	37	6	Carbonyl, oxymethyl–	12.8	37	6
Carboxylate, alkyl–	11.7	40	8	Carbonyl, alkyl, acyc–	12.8	84	29
Steroid, 3-oxo–	11.1	24	2	Carboxamide	12.1	97	37
Carboxamide (NR2)	10.0	46	12	Diene, 1,4–	11.8	25	2
Pyridine(H), 3-carbonyl–	9.9	25	3	Pyridine(H), 3-carbonyl–	10.5	26	3
Diene, 1,4–	9.7	22	2	Pyran(H), 4-oxy–	10.5	17	0
Benzene, 1-carbonyl–	9.6	65	22	Ether, alkoxymethyl–	9.7	16	0
Benzene, 1-alkyl-,2-aminomethyl–	9.5	1	1	Pyran(H), 3-hydroxy–	9.0	15	0
Sulfonamide, aryl–	9.2	29	5	Steroid	8.8	26	4

The number of positive (Np) and negative (Nn) compounds that contain each feature are listed.

descriptors were included in all the QSAR models together with the structural features, and they were among the key descriptors for the CYP2C9 substrate and CYP2D6 inhibitor models.

The larger susceptibility of CYP2C9 towards binding compounds containing hydrogen bond acceptors compared to CYP2D6,[9,20] and the smaller polar surface area for CYP2D6 active compounds[70] is clearly reflected in this analyses. The difference in lipophilicity among CYP2C9 and CYP2D6 active compounds, stated previously,[9,20] is not seen in our analysis.

3.3. Prediction of CYP activity for EINECS compounds

The developed QSAR models were used to screen 51,067 EINECS compounds for CYP activity (Table 8). 2848 compounds were predicted to be CYP2C9 substrates, which is 20% of the compounds within the domain of the developed model. For the CYP2D6 substrates the corresponding number was 1583 compounds (9% of the compounds within model domain). Similarly, 3877 and 4932 compounds were predicted as CYP2C9 and CYP2D6 inhibitors, respectively, which amounts to 17% and 21% of the compounds within domain of the QSAR models. 27% of the EINECS chemicals were predicted to be in structural and activity domain for the CYP2C9 substrate model, 35% for both the CYP2C9 inhibitor and CYP2D6 substrate model, and 47% for the CYP2D6 inhibitor model. According to these results, a substantial amount of existing chemicals can act either as substrates and/or inhibitors to these two major CYPs, and thus we can expect that many chemicals in consumer products and in the environment do affect human physiology through interactions with major metabolic pathways.

Nebert and Dalton stated in their review that 'Human CYP2D6 almost exclusively metabolizes drugs, and very few non-drug substrates'.[8] Our results on the other hand indicate that this isozyme indeed metabolizes many other chemicals. It would be peculiar if such an important enzyme with respect to drug metabolism, would not interact with other compounds that match the structural profile of the CYP2D6 binding pocket.

Rendic[62] estimated based on their collection of available clinical data that 25% of drugs act as substrates and 27% as inhibitors with members of the CYP-2C family. For the CYP2D6 isozyme they found that 15% interact as substrates and 22% as inhibitors. According to an investigation by Zanger et al.[9] conducted on 200 of the most sold drugs on the US market, 17% were identified to undergo oxidative metabolism via CYP2C9 and 15% via CYP2D6. Although, our screening results are within another chemical universe, we can conclude that our predictions for EINECS chemicals are within the same order of magnitude as the estimates given for drugs in the literature.

CYP2D6 is known to be a particularly important isozyme for metabolizing aromatic amines. To verify if this was reflected in the screening results, we examined the occurrences of aromatics amines among the EINECS chemicals screened with the CYP2D6 and CYP2C9 substrate and inhibitor model (Table 8). It was clearly seen that while 61% of all the EINECS compounds predicted as CYP2D6 substrates were classified as aromatic amines, the corresponding value for CYP2D6 inhibitors was only 18%. Around 30% of the CYP2D6 non-substrates and non-inhibitors were classified as aromatic amines, which corresponds to the percentage of aromatic amines within the whole EINECS list. The same was the case

Table 7
The best correlated positive and negative features for the CYP2D6 inhibitor sub-sets, evaluated by χ^2 statistics

Key features for 2D6 inhibitors, set a	χ^2	Nn	Np	Key features for 2D6 inhibitors, set b	χ^2	Nn	Np
tert-Amine, alkyl–	20.7	82	98	*tert*-Amine	29.7	72	98
seic-Amine(NH), alkyl–	19.0	15	36	Ethane, 1-alkylamino-,2-hydroxy–	14.0	4	18
sec-Amine(NH)	14.9	29	47	Amine, dimethyl, alkyl–	12.6	7	21
Benzene, 1-(alkyl, acyc)–	14.1	63	76	Benzene, 1-(alkyl, cyc)–	11.4	32	46
Alcohol, benzyl–	9.2	3	13	Pyridine(H), 2-(alkyl, acyc)–	10.9	5	17
Piperidine	9.2	27	39	*sec*-Amine(NH)	9.9	35	47
Piperidine, 2-(alkyl, acyc)–	9.2	0	8	Ethane, 1-amino-,2-hydroxy–	9.7	9	21
Benzene, 1-amino-,3-methoxy–	7.8	0	7	Oxycarbonyl, *O*-methyl–	8.8	4	14
Benzene, 1-(2-aminoethyl)–	6.4	19	28	Methane, 1-aryl-,1-phenyl–	8.1	1	9
Benzene, 1-alkyl-,4-chloro–	6.1	5	13	Piperidine	7.8	29	39
Methane, 1-alkylamino-,1-phenyl–	6.1	5	13	Benzene, 1-alkyl-,4-chloro–	7.6	4	13
Pyridine(H), 2-(alkyl, acyc)–	5.6	9	17	Benzene, 1-benzyl–	6.8	7	16
Benzene, 1-alkylaminomethyl–	5.5	10	18	Propane, 1-alkylamino-,3-arylamino–	6.7	1	8
Benzene, 1-alkyl-,2-methoxy–	4.9	0	5	Benzene, 1-alkoxy-,4-alkyl–	6.6	9	18
Benzene, 1-alkoxy-,3-alkylamino–	4.9	0	5	Benzene, 1-chloro–	6.4	30	38
Pyridine(H), 3-oxymethyl–	4.9	0	5	Azepine	6.0	2	9
Propane, 1-amino-,3-hydroxy–	4.9	0	5	Ethane, 1-hydroxy-,2-phenoxy–	6.0	2	9
Benzene, 1-aminomethyl(CH_2)–	4.8	7	14	Methane, 1-aryl-,1-hydroxy–	5.8	7	15
Ethane, 1-alkylamino-,2-amino–	4.8	2	8	*sec*-Amine(NH), alkyl–	5.6	29	36
Carbonyl, alkyl, cyc–	15.5	69	2	Carbonyl, alkyl, cyc–	17.4	72	2
Ketone, alkenyl–	13.9	30	3	Carboxy	14.7	41	7
Steroid	13.7	24	1	Ketone	13.2	56	15
Carbonyl, alkenyl, cyc–	10.8	48	13	Ketone, alkyl, cyc–	12.0	30	4
Ketone, alkyl, cyc–	10.6	28	4	Ketone, alkenyl–	10.9	26	3
Benzene, 1-alkyl-,4-halo–	8.6	9	2	Benzene, 1-alkyl-,4-halo–	9.9	8	2
Steroid, 3-oxo–	8.5	17	1	Sulfonamide, aryl–	9.5	24	3
Ketone	8.0	47	15	Steroid	9.1	18	1
Carboxamide	7.9	70	28	Carbonyl, alkenyl, cyc–	8.4	44	13
Benzene, 1-alkylamino-,2-alkyl–	7.3	2	1	Carboxamide	8.3	71	28
Steroid, 17-oxy–	7.0	15	1	Carboxamide (NR2)	7.9	33	8
Ether, phenyl–	6.8	43	5	Alkene, trisubst	7.8	37	1
Carboxamide (NR2)	6.2	30	8	1,3-Diazine(H), 4-oxo–	7.7	16	1
Ether, alkyl, cyc–	5.8	39	13	Benzene, 1-alkylamino-,2-alkyl–	7.3	2	1
Sulfonyl group	5.7	25	6	Steroid, 10-methyl–	7.0	12	0
Amine, *i*-propyl–	5.6	3	1	1-Benzene-sulfonamide	6.2	19	3
Cyclohexenone	4.9	12	1	Benzene, 1-(alkyl, acyc)–	5.9	78	76
Ketone, cyclopentyl–	4.9	9	0	Steroid, 17-oxy–	5.6	13	1

The number of positive (Np) and negative (Nn) compounds that contain each feature are listed.

for CYP2C9 active and inactive compounds, where all groups contained about 30% of aromatic amines. These results show that the overrepresentation of aromatic amines among the CYP2D6 substrates is clearly reflected by the predictions for the EINECS chemicals, which demonstrates that the QSAR model truly can identify a highly CYP2D6 relevant substrate class in a foreign data set.

It was seen from the EINECS screening results, that sulfonamides were a highly relevant compound class for both for CYP2C9 substrates and inhibitors. This was also reflected by the fact the sulfonamide containing fragments were seen among the most important positive fragments within the developed QSAR models. Thus sulfonamides (molecules containing a $S(=O)_2N$ group) were identified within the EINECS predictions by substructure search. In Table 8 it was seen that almost all sulfonamides were classified as both CYP2C9 substrates and inhibitors, and at the same time they were classified as CYP2D6 inactive. This example also illustrates the ability of the developed QSAR models to identify CYP2C9 active compound classes in a data set without available experimental data.

Poly-aromatic hydrocarbons were predicted to be slightly more likely to be CYP2C9 substrates or inhibitors rather the non-substrates or non-inhibitors. These compounds were predominantly classified as non-substrates and non-inhibitors to the CYP2D6 isozyme.

The overlap between EINECS compounds predicted as substrates and inhibitors for CYP2C9 and CYP2D6, respectively, was investigated (Fig. 2). The developed models thus predicted the overlap between CYP2C9 substrates and inhibitors to be 2.4–3 times larger than the number of compounds that were predicted to be exclusively CYP2C9 substrates or inhibitors. This was also illustrated by the similarities in the best correlated features for the inhibitor and the substrate models (Tables 4 and 5). The inhibitor data and thus also the models do not distinguish between reversible and irreversible inhibition,[3] and thus some of the overlap can be explained by compounds that both act as inhibitors and substrates as a part of the metabolism process. Much fewer EINECS compounds were predicted to be both substrates and inhibitors to the CYP2D6 isozyme, with a significantly larger portion of the chemicals predicted to only be inhibitors and significantly fewer to be only substrates.

It can thus be concluded that a noteworthy portion of the environmental chemicals within the EINECS data set were predicted to interact with CYP2C9 and CYP2D6, either to be metabolized by these isozymes or to inflict upon their activity through inhibition. As consumers are exposed to many of these compounds in larger and smaller quantities, their interactions with the CYP enzymes might affect metabolism of other compounds, like drugs, by competing for the same CYPs.

3.4. Example: Drug-herbal interactions and (S)-warfarin

Evidence for the interactions between the anticoagulant drug (S)-warfarin (Fig. 3) and several herbs used as food supplements has been described in the recent literature. It is relatively well established that patients using (S)-warfarin should avoid herbs like Danshen, Dong quai, *Ginko biloba* and St. John's worth, among others, due to increased risk of bleeding.[71] Several flavonoids found in

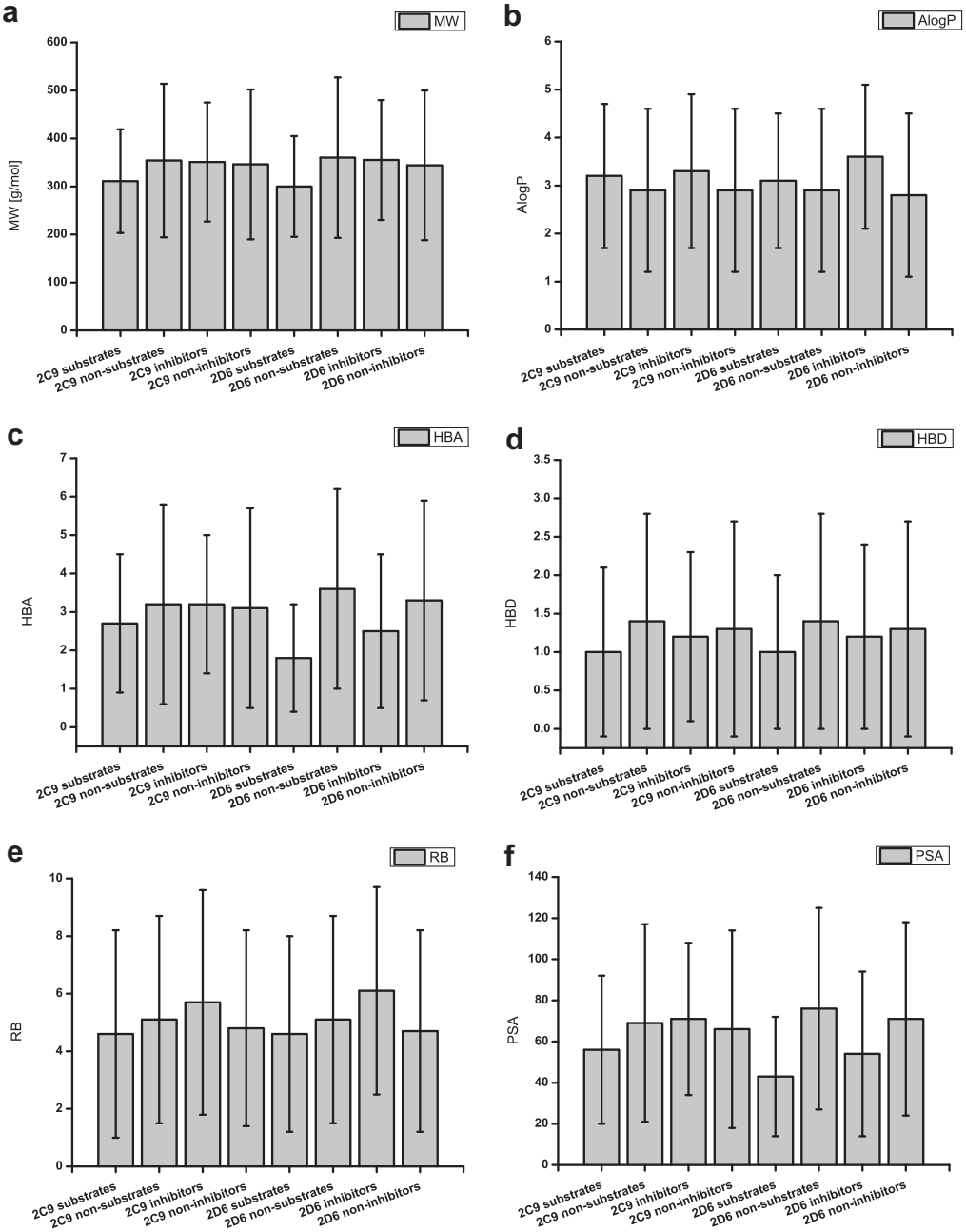

Figure 1. The mean values of the molecular weight (MW)(a), octanol–water partition coefficient (*A*log*P*)(b), number of hydrogen bond acceptors (HBA)(c) and donors (HBD)(d), no. rotatable bonds (RB)(e) and polar surface area (PSA)(f) for CYP substrates/non-substrates and CYP inhibitors/non-inhibitors in the full training sets used. The standard deviations are shown as thin lines.

Ginko biloba have been identified as CYP2C9 inhibitors, with the flavones amentoflavone, apigenin and luteolin having IC50 values below 10 μM.[71] CYP2C9 inhibitory activity was also documented for other groups of flavonoids,[56,72–74] although other flavonoids were found to have limited CYP2C9 activity.[75]

A study that used flexible docking and molecular dynamics simulations, showed that a group of flavones and flavonols, including apigenin and luteolin, have the ability to bind close to the heme and occupy the substrate binding site observed for flurbiprofen in CYP2C9.[55] In another study, flavone and five other flavonoid inhibitors of CYP2C9 could be docked into the substrate binding site.[56]

In the published crystal structure, (S)-warfarin was seen to bind in a position which is believed to be too far from the heme group in the substrate binding site of CYP2C9 for hydroxylation to occur. The authors thus speculated that (S)-warfarin would subsequently move closer to the heme group by conformational changes in order to undergo hydroxylation.[76] One can speculate that flavoids and other compounds that bind outside the active site of CYP2C9 and are considered to act as non-competitive CYP2C9 inhibitors, might interfere with the metabolism of (S)-warfarin.

We used our in-house database tool to search for all compounds containing a flavone group (Fig. 4) in our database, and explored the predicted CYP2C9 and CYP2D6 activity for these compounds. Of the 60 compounds containing a flavone substructure and within structural and activity domain of the CYP2C9 inhibitor QSAR model, all were predicted to act as CYP2C9 inhibitors. Fifty-three compounds containing a flavone group were within the model domain

Table 8
EINECS compounds predicted as CYP2D6 and CYP2C9 substrates, non-substrates, inhibitors and non-inhibitors, respectively, presented as counts and percentage of predictions within the domain of the QSAR models

Predicted class	CYP2D6 substrates	CYP2D6 non-substrates	CYP2C9 substrates	CYP2C9 non-substrates
Predictions within model domain	1583 (9%)	16,359 (91%)	2848 (20%)	11,066 (80%)
- aromatic amines	**973 (61%)**	5112 (31%)	927 (33%)	3467 (31%)
- sulfonamides	*10* (*0.6%*)	555 (3%)	**229 (8%)**	*24* (*0.2%*)
- poly-aromatic hydrocarbons	*18* (*1%*)	745 (5%)	113 (4%)	*261* (*2%*)
Predicted class	CYP2D6 inhibitors	CYP2D6 non-inhibitors	CYP2C9 inhibitors	CYP2C9 non-inhibitors
Predictions within model domain	4932 (21%)	18,997 (79%)	3877 (17%)	18,671 (83%)
- aromatic amines	*893* (*18%*)	6012 (32%)	988 (25%)	5102 (27%)
- sulfonamides	*33* (*0.7%*)	575 (3%)	**298 (8%)**	*31* (*0.2%*)
- poly-aromatic hydrocarbons	*51* (*1%*)	819 (4%)	165 (4%)	497 (3%)

In each case the percentage of aromatic amines, sulfonamides and poly-aromatic hydrocarbons among the predictions within domain, are listed. Compound classes that are significantly overrepresented compared to the full list of EINECS compounds are shown in bold and those which are under-represented in italics.

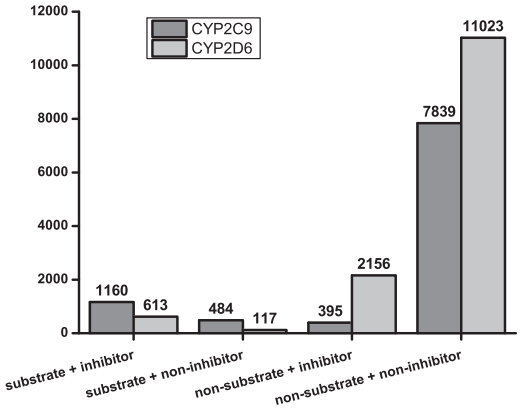

Figure 2. Number of EINECS compounds predicted both to be a substrate and an inhibitor, compared to the number of compounds that are predicted either to be only a substrate or only an inhibitor to CYP2C9 and CYP2D6, respectively. For each of the CYP isozymes, this analysis was made for compounds within the domain of both the substrate QSAR and the inhibitor QSAR. The number of compounds belonging to the different classes, are listed above the relevant columns.

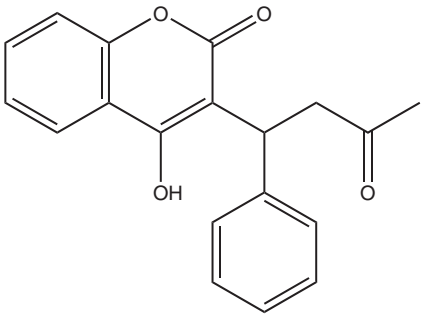

Figure 3. 2D representation of warfarin.

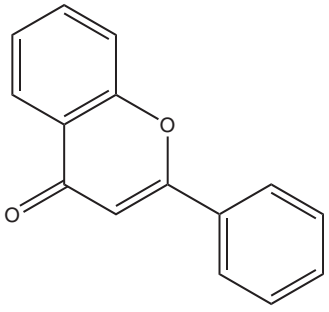

Figure 4. 2D representation of flavone.

of the CYP2C9 substrate model. Forty-three of the compounds within model domain were classified as CYP2C9 substrates, with 42 of those predicted to be inhibitors as well. Five of the ten substances that were predicted to be CYP2C9 non-substrates, were classified to be CYP2C9 inhibitors. Among the predicted CYP2C9 inhibitors, compounds like flavone and apigenin that have been experimentally identified as inhibitors are present, as well as many structures that have not yet been tested for CYP2C9 inhibition activity.

All the 65 compounds containing a flavone substructure and within domain of both the CYP2D6 substrate and the CYP2D6 inhibitor models, were predicted to be CYP2D6 inactive, that is both non-substrates and non-inhibitors. None of the substances within model domain were identified as CYP2D6 active.

Thus according to our predictions, compounds containing a flavone substructure were generally classified as CYP2C9 active and CYP2D6 inactive. These results illustrate that our QSAR models can be used to explore CYP activity for specific compounds, as well as compound classes, and potentially to identify risk of interactions with (S)-warfarin or other drugs with narrow therapeutic index. It could be particularly useful to identify such problems for new food and herbal products before they enter the market.

3.5. Analysis of procarcinogenic potential among CYP active compounds

Groups of compounds containing substructures that are characteristic of known procarcinogens were identified among compounds predicted to be CYP active, as well as predicted as either positive or negative in the Ames mutagenicity test (Table 9). As a positive test result in the Ames in vitro assay is a good marker for mutagenicity and for identifying mutagenic carcinogens, this model was chosen for the analysis.

We searched for simple substructures characterizing these compound classes; two inter-connected six-member aromatic rings for poly-aromatic hydrocarbons, a combination of a six-member aromatic ring and a tertiary nitrogen for aromatic amines, six-member and five-member aromatic rings containing one or two nitrogen atoms to search for the most common heterocyclic amines and the N–N=O group for nitrosamines.

While an overrepresentation of poly-aromatic hydrocarbons was seen for predicted Ames positive CYP2C9 substrates and inhibitors compared to predicted Ames positive CYP2C9 non-substrates and non-inhibitors, the CYP2D6 did not favor this compound class. Thus the modeling results support the finding that specific variant alleles of CYP2C9 have been linked to the metabolism of procarcinogens and increased risk of cancer.[8] The majority of all poly-aromatic hydrocarbons were identified as Ames positive according to the model predictions.

Table 9
The percentage of aromatic hydrocarbons and of four classes of compounds known to be rich in procarcinogens (poly-aromatic hydrocarbons, aromatic amines, heterocyclic amines and nitrosamines) with respect to predicted activity in the four CYP models and the Ames mutagenicity model

Model predictions by	CYP activity	Ames activity	Aromatic hydrocarb.	Poly-arom. hydrocarb.	Aromatic amines	Heterocycl. amines	Nitros amines
Ames model alone		POS	70	**15**	**54**	**19**	**2.6**
		NEG	52	1	31	9	0.0
CYP2C9 substrates	POS	POS	93	**22**	54	3	0.4
+Ames	POS	NEG	74	1	24	3	0.0
	NEG	POS	61	6	51	8	0.2
	NEG	NEG	36	1	18	3	0.0
CYP2C9 inhibitors	POS	POS	78	21	**61**	5	0.3
+Ames	POS	NEG	36	1	14	2	0.0
	NEG	POS	56	10	36	8	0.4
	NEG	NEG	39	1	18	4	0.0
CYP2D6 substrates	POS	POS	90	7	66	5	0.0
+Ames	POS	NEG	75	0	56	2	0.0
	NEG	POS	71	15	57	6	0.3
	NEG	NEG	37	1	17	3	0.0
CYP2D6 inhibitors	POS	POS	54	12	37	5	0.0
+Ames	POS	NEG	30	0	14	2	0.0
	NEG	POS	65	12	49	8	0.5
	NEG	NEG	41	1	18	5	0.0

Only predictions that are within the domain of both the CYP model considered and the Ames model are considered. Highly overrepresented contributions among the Ames positive predictions are marked with bold.

Aromatic amines were seen to be overrepresented among the compounds predicted as Ames positive compared to those predicted as Ames negatives. The percentages of the predicted Ames positive CYP active compounds were compared to the percentages of the corresponding Ames positive CYP inactive compounds. In most cases these were considered comparable, taking the uncertainties of the predicted data into account. An exception to this was the CYP2C9 inhibitors, for which the percentage of Ames positive compounds was twice as high, than for the non-inhibitors for this compound class. Although CYP2D6 is known to metabolize aromatic amines, this isozyme is generally not considered to contribute to procarcinogenic metabolism,[8] which is in agreement with the findings of our analysis.

Ames positive heterocyclic amines and nitrosamines cannot be considered to be favored by any of these CYPs, and nitrosamines did not interact with CYP2D6 at all according to our model predictions.

Our results thus showed some tendency towards increased CYP2C9 activity among poly-aromatic hydrocarbons, a compound class rich in procarcinogens, but no such tendencies were seen for the CYP2D6 active compounds.

4. Conclusions

Four new QSAR models for substrate recognition and inhibitor identification for CYP2C9 and CYP2D6 were developed based on human clinical data, with the use of structural features and molecular descriptors relevant with respect to ligand binding. Each of the models used a set of the best correlated structural features generated from the extensive library of Leadscope. Robust models were developed, with sensitivities ranging from 66% to 73% and specificities in the range 82–88%.

The models were used to screen 51,067 EINECS chemicals with respect to CYP activity. The chemicals within the applicability domain of the respective models were thus predicted to be either substrates or inhibitors, both substrates and inhibitors, or inactive to these two isozymes. Where the majority of CYP2C9 active compounds were predicted to be both a substrate and an inhibitor at the same time, the CYP2D6 active compounds were primarily predicted to be only inhibitors.

The structural characteristics of the models, reflected by the best correlated structural features, matched known binding properties of these isozymes. Substructure analysis among the EINECS compounds predicted as CYP substrates and inhibitors illustrate the good ability of the developed QSAR models to identify compound classes that act as substrates to the two CYPs, for example, sulfonamides for CYP2C9 and aromatic amines for CYP2D6.

Compounds containing a flavone substructure were generally predicted to be CYP2C9 inhibitors. This supports literature studies on the CYP2C9 inhibitory effect of flavone compounds that can potentially affect the activity of drugs like (S)-warfarin. Furthermore, a weak overrepresentation of poly-aromatic hydrocarbons was seen among predicted CYP2C9 active and Ames positive compounds, indicating that CYP2C9 plays a role in the metabolism of procarcinogens.

Acknowledgements

T.R. and N.G.N. would like to thank for financial support from the Danish Environmental Agency and from a Grant on ‘in vitro metabolism’ from the Danish in vitro network.

References and notes

1. Zhou, S. *Curr. Drug Metab.* **2008**, *9*, 310.
2. Plant, N. *Biochim. Biophys. Acta* **2007**, *1770*, 478.
3. Pelkonen, O.; Turpeinen, M.; Hakkola, J.; Honkakoski, P.; Hukkanen, J.; Raunio, H. *Arch. Toxicol.* **2008**, *82*, 667.
4. Guengerich, F. P. *Chem. Res. Toxicol.* **2008**, *21*, 70.
5. Wienkers, L. C.; Heath, T. G. *Nat. Rev. Drug Disc.* **2005**, *4*, 825.
6. Williams, J. A.; Hyland, R.; Jones, B. C.; Smith, D. A.; Hurst, S.; Goosen, T. C.; Peterkin, V.; Koup, J. R.; Ball, S. E. *Drug Metab. Dispos.* **2004**, *32*, 1201.
7. Kalgutkar, A. S.; Obach, R. S.; Maurer, T. S. *Curr. Drug Metab.* **2007**, *8*, 407.
8. Nebert, D. W.; Dalton, T. P. *Nat. Rev. Cancer* **2006**, *6*, 947.
9. Zanger, U. M.; Turpeinen, M.; Klein, K.; Schwab, M. *Anal. Bioanal. Chem.* **2008**, *392*, 1093.
10. Paine, M. F.; Hart, H. L.; Ludington, S. S.; Haining, R. L.; Rettie, A. E.; Zeldin, D. C. *Drug Metab. Dispos.* **2006**, *34*, 880.
11. Roy, K.; Roy, P. P. *Expert Opin. Drug Metab. Toxicol.* **2009**, *5*, 1245.
12. Shimada, T.; Yamazaki, H.; Mimura, M.; Inui, Y.; Guengerich, F. P. *J. Pharmacol. Exp. Ther.* **1994**, *270*, 414.
13. Ingelman-Sundberg, M.; Sim, S. C.; Gomez, A.; Rodriguez-Antona, C. *Pharmacol. Ther.* **2007**, *116*, 496.
14. DeLisle, R. K.; Otten, J.; Rhodes, S. *Comb. Chem. High Throughput Screen.* **2011**, *14*, 396.
15. Zanger, U. M.; Raimundo, S.; Eichelbaum, M. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **2004**, *369*, 23.
16. Ingelman-Sundberg, M. *Pharmacogenomics J.* **2005**, *5*, 6.
17. Poolsup, N.; Po, A. L. W.; Knight, T. L. *J. Clin. Pharm. Ther.* **2000**, *25*, 197.
18. Le Bourdonnec, B.; Leister, L. K. *Curr. Med. Chem.* **2009**, *16*, 3093.
19. Chen, Q.; Zhang, T.; Wang, J.; Wei, D. *Curr. Drug Metab.* **2011**, *12*, 436.
20. Anzenbacher, P.; Anzenbacherova, E. *Cell. Mol. Life Sci* **2001**, *58*, 737.
21. Ekins, S.; De Groot, M. J.; Jones, J. P. *Drug Metab. Dispos.* **2001**, *29*, 936.

22. de Groot, M. J.; Ekins, S. *Adv. Drug Deliv. Rev.* **2002**, *54*, 367.
23. Freitas, R. F.; Bauab, R. L.; Montanari, C. A. *J. Chem. Inf. Model.* **2010**, *50*, 97.
24. Zhou, S. F.; Zhou, Z. W.; Yang, L. P.; Cai, J. P. *Curr. Med. Chem.* **2009**, *16*, 3480.
25. Gautier, J. C.; Lecoeur, S.; Cosme, J.; Perret, A.; Urban, P.; Beaune, P.; Pompon, D. *Pharmacogenetics* **1996**, *6*, 489.
26. Shou, M.; Krausz, K. W.; Gonzalez, F. J.; Gelboin, H. V. *Arch. Biochem. Biophys.* **1996**, *328*, 201.
27. Faust, R. A toxicity summary for dibenz[*a*,*h*]antrahacene, Oak Ridge Reservation Environmental Restoration Program, 1995.
28. Refsgaard, H. H. F.; Jensen, B. F.; Christensen, I. T.; Hagen, N.; Brockhoff, P. B. *Drug Dev. Res.* **2006**, *67*, 417.
29. Ewing, T.; Feher, M. *Mol. Inform.* **2010**, *29*, 127.
30. Li, H.; Sun, J.; Fan, X.; Sui, X.; Zhang, L.; Wang, Y.; He, Z. *J. Comput. Aided Mol. Des.* **2008**, *22*, 843.
31. Burton, J.; Ijjaali, I.; Barberan, O.; Petitet, F.; Vercauteren, D. P.; Michel, A. *J. Med. Chem.* **2006**, *49*, 6231.
32. Yap, C. W.; Chen, Y. Z. *J. Chem. Inf. Model.* **2005**, *45*, 982.
33. Hudelson, M. G.; Jones, J. P. *J. Med. Chem.* **2006**, *49*, 4367.
34. Gleeson, M. P.; Davis, A. M.; Chohan, K. K.; Paine, S. W.; Boyer, S.; Gavaghan, C. L.; Arnby, C. H.; Kankkonen, C.; Albertson, N. *J. Comput. Aided Mol. Des.* **2007**, *21*, 559.
35. Hammann, F.; Gutmann, H.; Baumann, U.; Helma, C.; Drewe, J. *Mol. Pharm.* **2009**, *6*, 1920.
36. Mishra, N. K.; Agarwal, S.; Raghava, G. P. S. *BMC Pharmacol.* **2010**, *10*, 8.
37. Ringsted, T.; Nikolov, N.; Jensen, G. E.; Wedebye, E. B.; Niemela, J. *SAR QSAR Environ. Res.* **2009**, *20*, 309.
38. Kriegl, J. M.; Arnhold, T.; Beck, B.; Fox, T. *QSAR Comb. Sci.* **2005**, *24*, 491.
39. Jensen, B. F.; Vind, C.; Padkjaer, S. B.; Brockhoff, P. B.; Refsgaard, H. H. F. *J. Med. Chem.* **2007**, *50*, 501.
40. Dapkunas, J.; Sazonovasay, A.; Japertas, P. *Chem. Biodivers.* **2009**, *6*, 2101.
41. Byvatov, E.; Baringhaus, K.; Schneider, G.; Matter, H. *QSAR Comb. Sci.* **2007**, *26*, 618.
42. Michielan, L.; Terfloth, L.; Gasteiger, J.; Moro, S. *J. Chem. Inf. Model.* **2009**, *49*, 2588.
43. Kontijevskis, A.; Komorowski, J.; Wikberg, J. E. S. *J. Chem. Inf. Model.* **2008**, *48*, 1840.
44. Afzelius, L.; Zamora, I.; Ridderstrom, M.; Andersson, T.; Karlen, A.; Masimirembwa, C. M. *Mol. Pharmacol.* **2001**, *59*, 909.
45. Peng, C.; Rushmore, T.; Crouch, G. J.; Jones, J. P. *Bioorg. Med. Chem.* **2008**, *16*, 4064.
46. Vaz, R.; Nayeem, A.; Santone, K.; Chandrasena, G.; Gavai, A. V. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3816.
47. Rossato, G.; Ernst, B.; Smiesko, M.; Spreafico, M.; Vedani, A. *ChemMedChem* **2010**, *5*, 2088.
48. Saraceno, M.; Massarelli, I.; Imbriani, M.; James, T. L.; Bianucci, A. M. *Chem. Biol. Drug Des.* **2011**, *78*, 236.
49. Vaz, R. J.; Zamora, I.; Li, Y.; Reiling, S.; Shen, J.; Cruciani, G. *Expert Opin. Drug Metab. Toxicol.* **2010**, *6*, 851.
50. Tarcsay, A.; Keseru, G. M. *Expert Opin. Drug Metab. Toxicol.* **2011**, *7*, 299.
51. Mendieta-Wejebe, J. E.; Correa-Basurto, J.; Garcia-Segovia, E. M.; Ceballos-Cancino, G.; Rosales-Hernandez, M. C. *Curr. Drug Metab.* **2011**, *12*, 533.
52. Lewis, D. F. V.; Ito, Y. *Expert Opin. Drug Metab. Toxicol.* **2010**, *6*, 661.
53. de Graaf, C.; Vermeulen, N. P. E.; Feenstra, K. A. *J. Med. Chem.* **2005**, *48*, 2725.
54. Tarcsay, A.; Kiss, R.; Keseru, G. M. *J. Comput. Aided Mol. Des.* **2010**, *24*, 399.
55. Si, D.; Wang, Y.; Zhou, Y.; Guo, Y.; Wang, J.; Zhou, H.; Li, Z.; Fawcett, J. P. *Drug Metab. Dispos.* **2009**, *37*, 629.
56. Shimada, T.; Tanaka, K.; Takenaka, S.; Murayama, N.; Martin, M. V.; Foroozesh, M. K.; Yamazaki, H.; Guengerich, F. P.; Komori, M. *Chem. Res. Toxicol.* **2010**, *23*, 1921.
57. Unwalla, R. J.; Cross, J. B.; Salaniwal, S.; Shilling, A. D.; Leung, L.; Kao, J.; Humblet, C. *J. Comput. Aided Mol. Des.* **2010**, *24*, 237.
58. Moors, S. L. C.; Vos, A. M.; Cummings, M. D.; Van Vlijmen, H.; Ceulemans, A. *J. Med. Chem.* **2011**, *54*, 6098.
59. Li, J.; Schneebeli, S. T.; Bylund, J.; Farid, R.; Friesner, R. A. *J. Chem. Theory Comput.* **2011**, *7*, 3829.
60. Sheridan, R. P.; Korzekwa, K. R.; Torres, R. A.; Walker, M. J. *J. Med. Chem.* **2007**, *50*, 3173.
61. Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R. *J. Med. Chem.* **2005**, *48*, 6970.
62. Rendic, S. *Drug Metab. Rev.* **2002**, *34*, 83.
63. Roberts, G.; Myatt, G. J.; Johnson, W. P.; Cross, K. P.; Blower, P. E. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1302.
64. Campbell, I. *Stat. Med.* **2007**, *26*, 3661.
65. Official Journal of the European Union 1990, OJC 146 A.
66. Official Journal of the European Union 2002, OJC C 54/13.
67. Nikolov, N.; Grancharov, V.; Stoyanova, G.; Pavlov, T.; Mekenyan, O. *J. Chem. Inf. Model.* **2006**, *46*, 2537.
68. Niemelä, J.; Wedebye, E. B.; Nikolov, N. G.; Jensen, G. E.; Ringsted, T., The Advisory list for self-classification of dangerous substances, Environmental Project No. 1303 2010, Danish EPA, 2010.
69. Furge, L. L.; Guengerich, F. P. *Biochem. Mol. Biol. Edu.* **2006**, *34*, 66.
70. McMasters, D. R.; Torres, R. A.; Crathern, S. J.; Dooney, D. L.; Nachbar, R. B.; Sheridan, R. P.; Korzekwa, K. R. *J. Med. Chem.* **2007**, *50*, 3205.
71. Chavez, M. L.; Jordan, M. A.; Chavez, P. I. *Life Sci.* **2006**, *78*, 2146.
72. Kimura, Y.; Ito, H.; Ohnishi, R.; Hatano, T. *Food Chem. Toxicol.* **2010**, *48*, 429.
73. Moon, Y.; Kim, S. Y.; Ji, H. Y.; Kim, Y. K.; Chae, H.-J.; Chae, S.-W.; Lee, H. S. *Xenobiotica* **2007**, *37*, 246.
74. Quintieri, L.; Bortolozzo, S.; Stragliotto, S.; Moro, S.; Pavanetto, M.; Nassi, A.; Palatini, P.; Floreani, M. *Drug Metab. Pharmacokinet.* **2010**, *25*, 466.
75. Breinholt, V. M.; Offord, E. A.; Brouwer, C.; Nielsen, S. E.; Brosen, K.; Friedberg, T. *Food Chem. Toxicol.* **2002**, *40*, 609.
76. Williams, P. A.; Cosme, J.; Ward, A.; Angova, H. C.; Vinkovic, D. M.; Jhoti, H. *Nature* **2003**, *424*, 464.

[77]. Kirchheiner, J.; Brockmöller, J. *Clin. Pharmacol. Ther.* **2005**, *77*, 1.