

RESEARCH ARTICLE

# Predicting the Metabolic Sites by Flavin-Containing Monooxygenase on Drug Molecules Using SVM Classification on Computed Quantum Mechanics and Circular Fingerprints Molecular Descriptors

Chien-wei Fu, Thy-Hou Lin\*

Department of Pharmacy, National Taiwan University Hospital Hsin-Chu Branch, Institute of Molecular Medicine and Department of Life Science, National Tsing Hua University, HsinChu, Taiwan, ROC

\* [thlin@life.nthu.edu.tw](mailto:thlin@life.nthu.edu.tw)



**OPEN ACCESS**

**Citation:** Fu C-w, Lin T-H (2017) Predicting the Metabolic Sites by Flavin-Containing Monooxygenase on Drug Molecules Using SVM Classification on Computed Quantum Mechanics and Circular Fingerprints Molecular Descriptors. PLoS ONE 12(1): e0169910. doi:10.1371/journal.pone.0169910

**Editor:** Bin Liu, Harbin Institute of Technology Shenzhen Graduate School, CHINA

**Received:** July 29, 2016

**Accepted:** December 22, 2016

**Published:** January 10, 2017

**Copyright:** © 2017 Fu, Lin. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All the data generated in this work are attached as supplementary materials with the article and they are as follows: S1 Table, S2 Table, S3 Table, and S1 Figure.

**Funding:** This work has been supported by a grant MOST 105-2320-B007-001-MY2 from the Ministry of Science and Technology, Taiwan ROC. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Abstract

As an important enzyme in Phase I drug metabolism, the flavin-containing monooxygenase (FMO) also metabolizes some xenobiotics with soft nucleophiles. The site of metabolism (SOM) on a molecule is the site where the metabolic reaction is exerted by an enzyme. Accurate prediction of SOMs on drug molecules will assist the search for drug leads during the optimization process. Here, some quantum mechanics features such as the condensed Fukui function and attributes from circular fingerprints (called Molprint2D) are computed and classified using the support vector machine (SVM) for predicting some potential SOMs on a series of drugs that can be metabolized by FMO enzymes. The condensed Fukui function  $f_A^-$  representing the nucleophilicity of central atom A and the attributes from circular fingerprints accounting the influence of neighbors on the central atom. The total number of FMO substrates and non-substrates collected in the study is 85 and they are equally divided into the training and test sets with each carrying roughly the same number of potential SOMs. However, only N-oxidation and S-oxidation features were considered in the prediction since the available C-oxidation data was scarce. In the training process, the LibSVM package of WEKA package and the option of 10-fold cross validation are employed. The prediction performance on the test set evaluated by accuracy, Matthews correlation coefficient and area under ROC curve computed are 0.829, 0.659, and 0.877 respectively. This work reveals that the SVM model built can accurately predict the potential SOMs for drug molecules that are metabolizable by the FMO enzymes.

## Introduction

The flavin-containing monooxygenase (FMO) is a flavoprotein which carries a flavin adenine dinucleotide (FAD) and utilizes NADPH and oxygen to catalyze the metabolism of many xenobiotics such as compounds containing nitrogen, sulfur, selenium, phosphorous and other

**Competing Interests:** The authors have declared that no competing interests exist.

nucleophilic heteroatoms [1–4]. The family of mammalian FMO genes is comprised with five similar genes from FMO1 to FMO5 and all of them are important Phase I metabolic enzymes as being capable of metabolizing xenobiotics. FMO1 and FMO3 are the two major isoforms expressed in liver microsomes and other tissues. While FMO1 is highly expressed in fetal liver, FMO3 is predominantly found in adult human. However, FMO2 is expressed overwhelmingly in lung and fewer FMO4 and FMO5 isoforms are found in human body [3, 5]. People may suffer the so called "fish odor syndrome" when their FMOs are mutated or defected and failing to metabolize trimethylamine such as trimethylamine *N*-oxide to its oxygenated form for converting it into urine and sweat [6].

Both FMOs and cytochrome P450 (CYP450) are Phase I metabolic enzymes involving in the metabolism of xenobiotics in human. The main function of these microsomal enzymes is to add oxygen to the foreign compounds and render these foreign compounds to soluble form so that they can be excreted out of human body. Although both FMOs and CYP450 sometimes metabolize the same type of chemical compounds, the metabolic mechanisms exerted by these two enzymes are fundamentally different. While the CYP450 enzymes oxidize compounds through some electrophilic reactions and create some radical intermediates, the FMO enzymes oxidize compounds through the nucleophilic addition reactions [5, 7]. The substrates of the two enzymes are also quite different. The FMO enzymes tend to metabolize molecules with soft nucleophilic atoms such as N, S, P, and Se, whereas the CYP450 systems not only oxidize molecules with these types of atoms but also directly metabolize the C atom and few of this type is found by FMO enzymes.

Though the substrate range of FMO appear to be somewhat narrower than that of CYP450, the enzyme system still plays a crucial role in Phase I metabolism of xenobiotics [7, 8]. For examples, some therapeutic agents such as benzydamine, itopride, and arbidol are primarily metabolized by FMO but not by CYP450 enzymes [9–11]. However, some are both metabolized by FMO and CYP450 enzymes like diphenhydramine and ziprasidone [12–14]. Moreover, some metabolites of FMO enzymes are also known to be the substrates of CYP450 such as TG100435, a Src kinase inhibitor [15]. The paramount difference between FMO and CYP450 is that unlike the latter, the former is not easily induced nor readily inhibited thereby the potential adverse drug-drug interaction may be prevented for drugs predominantly metabolized by the former [7, 8]. These properties may offer advantages in drug design and discovery. By taking the FMO detoxication pathways into account in designing new drugs, more meaningful drug-like materials may be found.

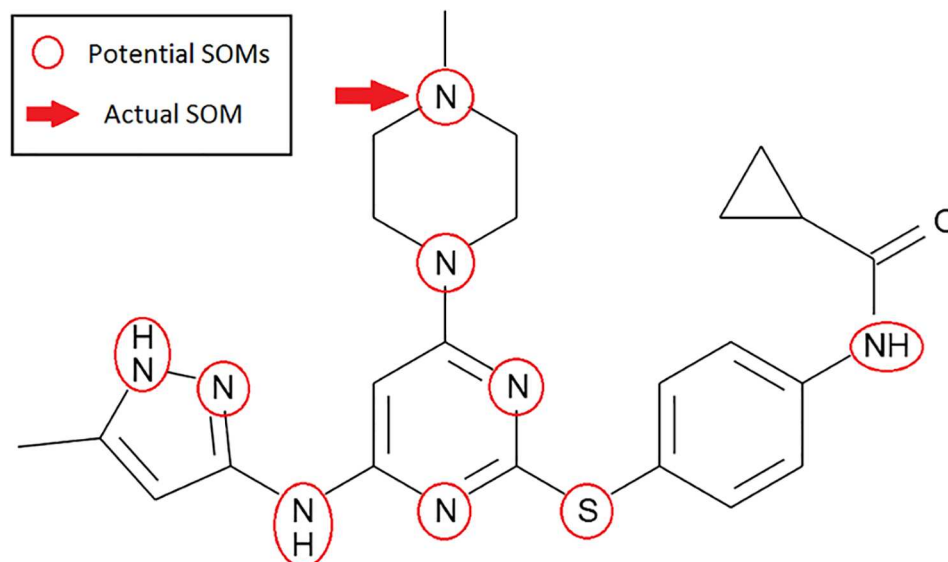
The site of metabolism (SOM) in a molecule is defined as the place where the reaction of metabolism can take place. Identification of SOMs is important in the drug optimization process for searching the drug leads. The information of SOM identified may guide developers to optimize the drug structure through changing the specific sites to avoid the unwanted metabolic reactions. However, identifying SOM through experimental procedures is not an easy task since each metabolite of synthesized compound has to be isolated and then characterized through some specific techniques such as LC-MS-MS [16]. Thus, acquiring information of SOM through experimental works is a highly cost and time consuming process. Therefore, developing accurate in silico methods for predicting SOM may be worthwhile and important.

There are several in silico methods for predicting SOM for drugs that can be metabolized by CYP450 have been developed. These approaches can be roughly divided into the following two categories namely, the ligand (substrate) based and the structure (enzyme) based ones. For examples, SMARTCyp [17] is a method that using a database of activation energies computed from quantum mechanics for a variety of ligand fragments and an accessibility descriptor computed to rank all the possible SOMs. IMPACTS [18] combines the docking result given by the molecular docking program GOLD, a reactivity index with computed hydrogen bond order

descriptors and the local ionization energy to refine the prediction for SOMs [19, 20]. Metasite combines the ligand and structure modeling together to predict SOM and the FMO enzymes are recruited into their package recently [21]. There are also machine learning techniques being developed for predicting SOM. MetaPrint2D [22] is an online tool that is based on training the Accelrys Metabolite Database [23] and then predicting SOM through counting occurrences of atomic fingerprints exhibiting as SOM or non-SOM in the database. Another example is RegioSelectivity (RS)-predictor [24, 25] which uses a Support Vector Machine (SVM) to predict SOM through training on 148 topological and 392 quantum mechanics atomic descriptors including contributions from neighbouring atoms.

There are other machine learning approaches using different descriptors to predict SOM such as Xenosite [26] using the Daylight fingerprint descriptors [27]. A SVM method [28] using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database [29] and other various descriptors has been published. FAME [30] is a broad metabolism predictor using the Random Forest model and the atomic descriptors calculated from CDK [31]. A method using the decision trees on the activation energies and solvent accessible surface area calculated from MOE [32–35] has been developed. Recently, the PASS algorithm using some 2D fingerprint descriptors [36] or the RASCAL (Random Attribute Subsampling Classification) algorithm [37] using some circular fingerprints has been employed to predict SOM by CYP450 enzymes [38, 39].

In this study, we predict SOM by FMO enzymes through using SVM on some quantum mechanics and a circular fingerprints method Molprint2D. Only the substrates-based information is used here to predict SOMs by FMO enzymes. The potential SOMs on some FMO substrates are marked first. These are usually atom N and S which account for over 90% presence in FMO metabolites. An example of a marked FMO substrate is shown in Fig 1. Next, the local reactivity of these potential SOMs is characterized through the quantum mechanics calculation. These quantum mechanics plus attributes from Molprint2D fingerprints computed are then classified by SVM to predict whether a site could be a SOM or not. The prediction



**Fig 1. An example of FMO substrate and its SOMs.** The potential SOMs identified for Tozasertib which is one of the FMO substrates studied. The potential SOMs are marked with red circles while the actual SOM is highlighted by a red arrow.

doi:10.1371/journal.pone.0169910.g001

**Table 1. Dividing the 85 compounds collected into the training and the test sets.**

Number	compound name	reaction	Potential som	som or nonsom
1	(S)-nicotine	N OX	(S)-nicotine 3 N	nonsom
			(S)-nicotine 7 N	som
2	clozapine	N OX	clozapine 3 N	nonsom
			clozapine 4 N	nonsom
			clozapine 10 N	nonsom
			clozapine 18 N	som
3	cysteamine	S OX	cysteamine 3 N	nonsom
			cysteamine 4 S	som
4	disulfoton	S OX	disulfoton 2 S	nonsom
			disulfoton 5 S	nonsom
			disulfoton 12 S	som
...	...	...	...	...

↓Assigned by Research Randomizer

**Compound 2, 3, 9, 13, 14, 18, 22, 23, 24, 26, 27, 28, 30, 31, 32, 33, 35, 37, 39, 40, 41, 46, 47, 49, 50, 52, 54, 57, 59, 60, 61, 64, 66, 68, 70, 72, 74, 75, 77, 79, 82, 85 were assigned as the test set**

First, the total collected compounds was arranged and numbered in order of collection time. Then, 42 out of 85 compounds were assigned as the test set by the web server Research Randomizer and the rest were treated as the training set.

doi:10.1371/journal.pone.0169910.t001

accuracy of our SVM models are also validated using both internal and external data sets. While SVM has been successfully used for predicting SOMs for some substrates of CY450 enzymes, it has not been implemented for predicting SOMs for substrates of only FMO enzymes like the work presented herein.

## Materials and Methods

### Datasets

Unlike the CYP450 system, the available FMO substrates are far less. For example, only eight substrates of FMO3 in the Human Metabolome Database (HMDB) are listed [40]. However, we have collected a total of 85 compounds from literatures [11, 13, 15, 21, 41–86] for making a dataset of 228 potential SOMs. The 85 compounds collected including 73 FMO substrates and 12 non FMO substrates. These 228 instances were randomly divided into a training and a test set using the Research Randomizer serve. The number of compounds allocated in the training and the test set were respectively 42 and 43 (Table 1). Moreover, the number of SOMs and non-SOMs instances identified in each set was shown in Table 2. Only two major FMO-catalyzed reactions namely N-oxidation and S-oxidation were considered in this study since the data for C-oxidation was scarce. The actual SOMs or non-SOMs on each compound were visually defined and marked. After dividing the dataset into the training and the test sets, the number of instances counted for the former and the latter were 111 and 117, respectively.

**Table 2. The total instances assigned for the training and the test sets were shown.**

Dataset	Number of compounds	Substrates of FMO	Non substrates of FMO	Number of instances (Potential SOMs)	Number of actual SOMs	Number of non-SOMs
The data for training with 10 CV	43	40	3	111	44	67
Test set	42	33	9	117	36	81

doi:10.1371/journal.pone.0169910.t002

## Quantum mechanics features

We employed the condensed Fukui reactivity indices as the main quantum mechanics features to represent the local reactivity of atoms within a molecule since molecules with larger values of Fukui function computed may show higher reactivity [87–92]. The Fukui function was defined as follows:

$$f(\vec{r}) = \left( \frac{\partial \rho(\vec{r})}{\partial N} \right)_v \quad (1)$$

Where  $\rho(\vec{r})$  was electronic density,  $N$  was number of electrons and  $v$  was external potential exerted by the nuclei. The concept was first described by Fukui in 1952 [93] and a corresponding definition with the Density functional theory (DFT) was given in 1984 [89, 91]. However, the condensed Fukui function was restricted to an atom within a molecule rather than a point in 3D space [94–98]. The condensed Fukui function or the Fukui reactivity indices of atom  $A$  in a molecule  $M$  were defined as follows:

$$f_A^+ = P_A(N + 1) - P_A(N) \quad (2)$$

$$f_A^- = P_A(N) - P_A(N - 1) \quad (3)$$

$$f_A^0 = 1/2 [P_A(N + 1) - P_A(N - 1)] \quad (4)$$

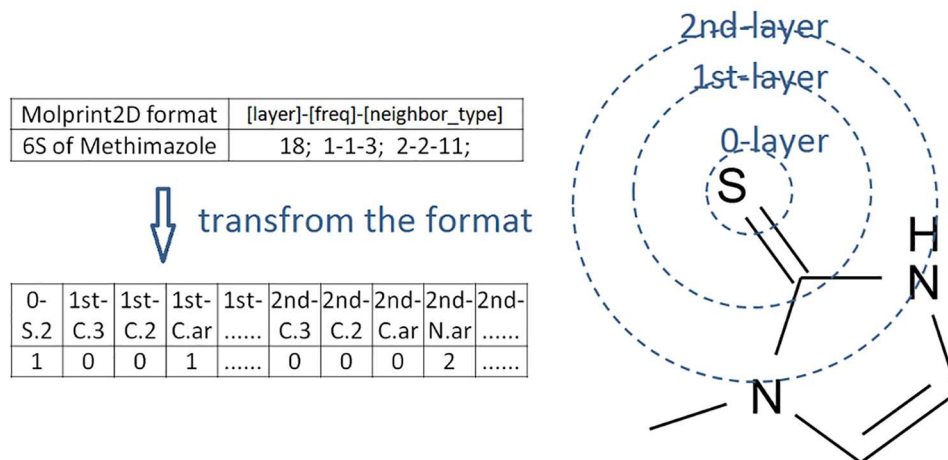
where  $f_A^+$  was the electrophilicity of atom  $A$ ,  $f_A^-$  was the nucleophilicity of atom  $A$ ,  $f_A^0$  was the radical attack susceptibility of atom  $A$ ,  $P_A(N)$  was the population on atom  $A$  with  $N$  electrons,  $P_A(N+1)$  was the population on atom  $A$  with  $N+1$  electrons, and  $P_A(N-1)$  was the population on atom  $A$  with  $N-1$  electrons. While  $P_A(N)$  was computed from the Mulliken charges,  $P_A(N)$  was computed as atomic number of atom  $A - q_A(N)$ , where  $q_A(N)$  was the charge on atom  $A$  with  $N$  electrons. The structures of all the 85 compounds were optimized in gas phase using the hybrid B3LYP functional and the 6-31G(d,p) basis set [99–103]. Then, the three population states with  $N$ ,  $N+1$  and  $N-1$  electrons were calculated using the optimized structures with the same basis set. The PCM solvation model was subsequently employed to calculate the surface area for each atom. The charges, parameters of condensed Fukui function, and surface area of atoms computed were used in the training process. All the aforementioned calculations were performed using the Gaussian 09 package [104].

## Attributes from circular fingerprints

The Fukui reactivity indices described above represented the local reactivity of a specific atom in a molecule and was insufficient to account the influence from neighbor atoms. To include the neighbor influence, we used Molprint2D [105] to evaluate the effect of neighbor atoms around each potential SOM. Molprint2D recognized the SYBYL atom type and counted the occurrence times of a neighbor atom of a particular atom type. The occurrence times counted for different neighbor atoms of different atom types were treated as the molecular descriptors. For a specific atom, all its neighbor atoms were generated iteratively by chemical bond lengths defined as follows:

[atomtype];  
 [1st-layer]-[frequency]-[neighbour\_type]; [2nd-layer]-[frequency]-[neighbour\_type]; [3rd-layer]-[frequency]-[neighbour\_type]; . . . ;

where [atomtype] was the atom type of a center atom designated by a number representing the atom type identified from all the 53 SYBYL atom types namely from C.3, C.2, C.ar, . . . , to



**Fig 2. An example of attributes generated by Molprint2D and its definition.** The original format of Molprint2D has been transformed by us into some numerical values for being readable by libSVM in the WEKA package.

doi:10.1371/journal.pone.0169910.g002

Co.oh, [1st-layer]-[frequency]-[neighbour\_type] represented its neighbor atom types within one bond, [2nd-layer]-[frequency]-[neighbour\_type] represented its neighbor atom types within two bonds and so on. An example of attributes from Molprint2D fingerprints was depicted in Fig 2. The Molprint2D fingerprints computed for each drug molecule were converted to a series of numeric attributes for being readable by a machine learning package. Here, we restricted the scope of each Molprint2D fingerprint computed to two chemical bond lengths. Further, the Molprint2D fingerprint for each drug molecule was generated using the Open Babel package [106].

### SVM classifier

The Support Vector Machine (SVM) classifier was used to predict whether an atom within a molecule could be a SOM of FMO enzymes or not. SVM can classify complex, non linear and high dimensional data into two classes. The merit of SVMs is to classify data by mapping input vectors into a high- or infinite- dimensional space with some kernel functions and then constructing a hyperplane or set of hyperplanes to separate them into two classes with a possible maximal margin computed. The margin is defined as the distance from the separating hyperplane to the nearest training-data point. The trained model of a SVM classifier can be used to predict to which class an unknown sample is belonging. Details on the basic SVM theory can be found elsewhere [107–109]. Here, we used the free SVM program LibSVM in the WEKA package [110] to perform all the classification tasks. Much of the running parameters used were default settings except C and  $\gamma$  values. We have selected the AUC based settings for training and then the best parameters obtained after training were used for predicting for the test set. Parameter optimization was performed by the package automatically using a grid-based search procedure.

### Performance measures

There were several model validation methods such as independent data test, n-fold cross-validation, jackknife (leave-one-out) cross-validation could be used to estimate the predictability of our SVM models built [111–114]. Due to fewer features (42 features) and samples (85 compounds) used, we have employed both the 10-fold cross-validation and jackknife (leave-one-

out) methods for validating our SVM models built. A web available method, the Research Randomizer was used to rationally divided our samples into a training and a test set. To avoid arbitrariness or bias, the SVM models built by the training set were validated not only by the 10-fold cross validation but also by the jackknife methods.

The prediction performance of the SVM models constructed was assessed through computing the following parameters: sensitivity (SE), specificity (SP), accuracy (ACC), and Matthews correlation coefficient (MCC) on the test set. These parameters were defined as follows:

$$\text{Sensitivity(SE)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

$$\text{Specificity(SP)} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (6)$$

$$\text{Accuracy(ACC)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (7)$$

$$\text{MCC} = \frac{(\text{TP})(\text{TN}) - (\text{FP})(\text{FN})}{\sqrt{[\text{TP} + \text{FP}][\text{TP} + \text{FN}][\text{TN} + \text{FP}][\text{TN} + \text{FN}]}} \quad (8)$$

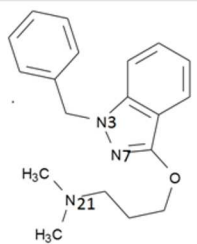
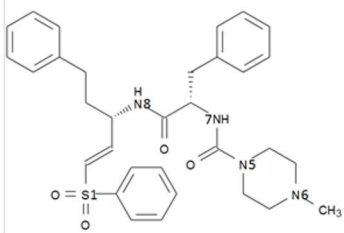
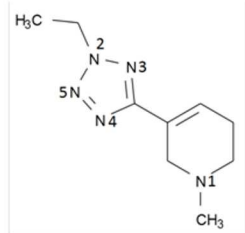
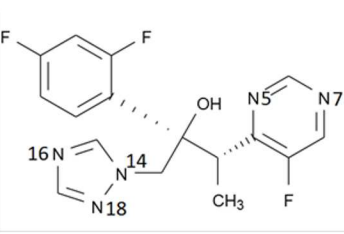
where TP, FN, TN, and FP stood for true positive, false negative, true negative and false positive, respectively. In addition to ACC and MCC, a receiver operating characteristic (ROC) analysis was also performed to evaluate the models. The ROC curves were constructed by plotting the false positive (1-SP) against the true positive rate (SE) computed and the area under a ROC curve (AUC) revealed whether a model constructed was a random model with an area of 0.5 or an ideal one with an area of 1.0 computed.

## Results and Discussion

### Condensed Fukui function

The total number of compounds studied was 85 and which were divided into a training and test set with each carrying roughly the same number of potential SOMs (Table 1). Both the singlet and triplet spin types of each optimized structure were considered in the calculation. The populations  $P_A(N)$ ,  $P_A(N+1)$ , and  $P_A(N-1)$  computed from Mulliken charges via DFT method were used to calculate the condensed Fukui Function. The values of  $f_A^+$ ,  $f_A^-$ , and  $f_A^0$  computed represent the electrophilicity of atom A, nucleophilicity of atom A, and radical attack susceptibility of atom A, respectively. The FMO enzymes are capable of metabolizing the xenobiotics carrying nucleophilic heteroatoms such as nitrogen and sulfur so that the corresponding  $f_A^-$  computed could be used to predict SOMs straightforwardly. In Fig 3, we present the values of condensed Fukui Function computed for 4 out of 85 compounds in the dataset. Among these 4 substrates, 3 carrying some N-oxidation SOMs and the last one carrying no SOM since it is a non-substrate of FMO enzymes (Fig 3).

The first drug shown in Fig 3 is benzydamine which is a nonsteroidal anti-inflammatory drug and usually used for pain relief and anti-inflammatory treatment of mouth and throat. There were three nitrogen atoms N3, N7, and N21 on benzydamine being predicted to be three potential SOMs and the corresponding  $f_A^-$  computed were 0.051, 0.045, and 0.050, respectively (Fig 3). However, N21 was the actual SOM of the drug though both its  $f_A^-$  and charge were not among the highest computed for the three potential SOMs. The actual SOM did not occur at the atom with the highest  $f_A^-$  computed which could be due to the substructure preference by FMO enzymes.

Structure	Instance	som or non	$q_A(N)$	$q_A(N+1)$	$q_A(N-1)$	vdw surface	$P_A(N)$	$P_A(N+1)$	$P_A(N-1)$	$f_A^+$	$f_A^-$	$f_A^0$
	benzydamin N3	nonsom	-0.43	-0.428	-0.379	2.37	7.43	7.43	7.38	-0.002	0.051	0.024
	benzydamin N7	nonsom	-0.374	-0.441	-0.329	10.4	7.37	7.44	7.33	0.067	0.045	0.056
	benzydamin N21	som	-0.407	-0.405	-0.358	5.77	7.41	7.41	7.36	-0.002	0.050	0.024
	K-11777 S1	nonsom	1.12	1.065	1.128	3.58	14.9	14.9	14.9	0.058	0.006	0.032
	K-11777 N6	nonsom	-0.504	-0.507	-0.498	2.87	7.50	7.51	7.5	0.003	0.007	0.005
	K-11777 N7	som	-0.421	-0.419	-0.373	5.57	7.42	7.42	7.37	-0.002	0.048	0.023
	K-11777 N8	nonsom	-0.61	-0.6	-0.6	2.12	7.61	7.6	7.6	-0.010	0.011	0.000
	K-11777 N9	nonsom	-0.528	-0.519	-0.522	3.91	7.53	7.52	7.52	-0.009	0.006	-0.002
	alvamine N1	nonsom	-0.433	-0.427	-0.3	5.58	7.43	7.43	7.3	-0.007	0.133	0.063
	alvamine N2	nonsom	-0.063	-0.092	-0.047	3.29	7.06	7.09	7.05	0.030	0.015	0.023
	alvamine N3	nonsom	-0.366	-0.491	-0.321	13.6	7.37	7.49	7.32	0.125	0.045	0.085
	alvamine N4	nonsom	-0.383	-0.438	-0.352	15.0	7.38	7.44	7.35	0.056	0.031	0.043
	alvamine N5	nonsom	-0.053	-0.164	0.0049	18.2	7.05	7.16	6.6	0.110	0.058	0.084
	voriconazole N5	nonsom	-0.449	-0.522	-0.427	11.8	7.45	7.52	7.48	0.072	0.023	0.048
	voriconazole N7	som	-0.4	-0.494	-0.358	16.5	7.4	7.49	7.36	0.094	0.042	0.068
	voriconazole N14	nonsom	-0.277	-0.272	-0.28	2.34	7.28	7.27	7.28	-0.005	-0.002	-0.004
	voriconazole N16	nonsom	-0.46	-0.481	-0.419	18.2	7.46	7.48	7.42	0.022	0.040	0.031
	voriconazole N18	nonsom	-0.262	-0.268	-0.195	15.4	7.26	7.27	7.2	0.006	0.067	0.037

**Fig 3. The values of condensed Fukui functions computed for the four selected examples in the training set.**  $q_A(N)$ ,  $q_A(N+1)$  and  $q_A(N-1)$  represent respectively the atomic charge in the molecule with N electrons, the atomic charge in the molecule with N+1 electrons, and the atomic charge in the molecule with N-1 electrons.  $P_A(N)$  is equal to the atomic number of atom A— $q_A(N)$  and so on.  $f_A^+$ ,  $f_A^-$  and  $f_A^0$  are the values of condensed Fukui function computed. The values of  $f_A^+$ ,  $f_A^-$  and  $f_A^0$  represent the electrophilicity of atom A, nucleophilicity of atom A, and radical attack susceptibility of atom A, respectively.

doi:10.1371/journal.pone.0169910.g003

There were five potential SOMs including both sulfur and nitrogen atoms S1, N6, N7, N8, and N9 on drug K11777 or k777 (N-methyl-piperazine-Phe-homoPhe-vinylsulfone-phenyl) which is an irreversible cysteine protease inhibitor (Fig 3). The corresponding  $f_A^-$  computed were 0.006, 0.007, 0.048, 0.011 and 0.006, respectively (Fig 3). The actual SOM of K11777 was on N7, a tertiary amine atom which also had the highest  $f_A^-$  computed among all the potential SOMs identified. However, the charge on N7 was among the second highest as opposite to the highest computed on S1 located in substructure  $R_2SO_2$ .

Voriconazole is a triazole antifungal drug usually used to treat the serious and invasive fungal infections. Apparently, voriconazole had five potential SOMs and they were on N5, N7, N14, N16 and N18, respectively (Fig 3). While N5 and N7 were located in substructure



pyrimidine, a six-members ring; N14, N16, and N18 were located in substructure triazole which was a five-members ring. The  $f_A^-$  values of N5, N7, N14, N16 and N18 computed were 0.023, 0.042, -0.002, 0.040 and 0.067, respectively. Note that the actual SOM in voriconazole occurred on N7 and its corresponding  $f_A^-$  was the second whereas its charge was the third highest computed among the five potential SOMs identified.

The fourth drug listed in Fig 3 was alvameline which was a M1 receptor agonist and M2/M3 receptor antagonist still under investigation for treating the Alzheimer's disease. The drug was a non-substrate of FMO enzymes and research attempt for the drug was currently ceased due to its poor clinical outcome. Although the drug was a non-substrate of FMO enzymes and no actual SOMs was given, there were five potential SOMs could be seen on N1, N2, N3, N4 and N5 (Fig 3). While N2, N3, N4 and N5 were in substructure tetrazole, a five-members ring, N1 was belonging to the tertiary amine of a six-members ring. The corresponding  $f_A^-$  computed for N1, N2, N3, N4 and N5 were 0.133, 0.015, 0.045, 0.031 and 0.058, respectively (Fig 3).

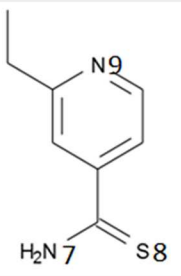
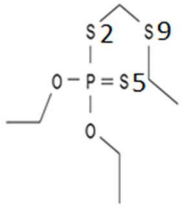
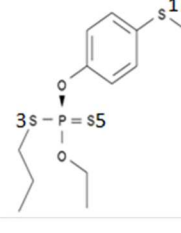
In general, the values of  $f_A^-$  computed representing the nucleophilicity of atom A in a drug molecule. We have found that the intramolecular ranking of  $f_A^-$  computed was highly correlated with the actual SOMs identified for each drug molecule. All these actual SOMs identified were either having the first or second ranked  $f_A^-$  computed. For examples, the  $f_A^-$  computed for the actual SOM of K11777 was ranked the highest, while those computed for benzydamine and voriconazole were both ranked the second highest among all the potential SOMs (Fig 3). This would indicate that all these identified actual SOMs had higher nucleophilicity for being recognizable by FMO enzymes. However, as shown in Fig 3, the intramolecular ranking of charge was not as effective as the aforementioned ranking of  $f_A^-$  in reflecting the actual SOMs identified. For examples, though the charges for the actual SOMs identified for benzydamine and K11777 were both secondly but not first ranked (Fig 3). In general, more negative charge would mean more nucleophilicity computed. However, unlike  $f_A^-$  computed, the charges computed were less effective or could not be used to correctly predict the actual SOMs. Here, both substrates and non-substrates of FMO enzymes were recruited into the training set for building a SVM model. We have found that some higher  $f_A^-$  values were computed within the non-substrate molecules. This would imply that simply using the condensed Fukui function computed was insufficient to predict whether a drug could be a substrate or non-substrate of FMO enzymes.

### Attributes from circular fingerprints

We employed the circular fingerprints of Molprint2D [105] to account the influence of surrounding atoms on a particular potential SOM. The format of Molprint2D generated using the 53 SYBYL atomic types C.3, C.2, C.ar, C.1, . . . , Mo, Mn, and Co.oh was converted into some 53×3 numerical attributes where the first, second, and third column was used to describe the center, first-layer and second-layer neighbors, respectively. For an example, a Molprint2D of 1; 1-1-2; 2-1-9; could be transformed into the first column represented by 53 attributes as 1 0 0 0 . . . . . 0, the second column represented by 53 attributes as 0 1 0 0 . . . . . 0, and the third column represented by 53 attributes as 0 0 0 0 0 0 0 1 0 0 0 . . . . . 0. Fig 4 gave the conversion of Molprint2D fingerprints generated into some numerical attributes for three selected drugs. Note that the columns with all zero values assigned for the hydrogen, sodium, . . . , and dummy atoms were removed.

### Features and model selection

The initial features employed were consisted by two parts: (1) 10 quantum mechanics features including charges, surface area, and condensed Fukui function computed; and (2) 53 × 3 attributes from layer 0 (central atom) to layer 2 created.

Structure	Potential SOM	Molprint2D format	S.3	N.ar	S.2	N.am	I-C.3	I- C.2	I- C.ar	I- P.3	II- C.3	II- C.ar	II- O.3	II- S.3	II- S.2	II-N.am	som or non
	Ethionamide N7	28; 1-1-2; 2-1-3;2-1-18;	0	0	0	1	0	1	0	0	0	1	0	0	1	0	nonsom
	Ethionamide N9	11; 1-2-3; 2-1-1;2-2-3;	0	1	0	0	0	0	2	0	1	2	0	0	0	0	nonsom
	Ethionamide S8	18; 1-1-2; 2-1-3;2-1-28;	0	0	1	0	0	1	0	0	0	1	0	0	0	1	som
	Phorate S2	10; 1-1-1;1-1-12; 2-2-8;2-2-10;	1	0	0	0	1	0	0	1	0	0	2	2	0	0	nonsom
	Phorate S5	10; 1-1-12; 2-2-8;2-1-10;	1	0	0	0	0	0	0	1	0	0	2	1	0	0	nonsom
	Phorate S9	10; 1-2-1; 2-1-1;2-1-10;	1	0	0	0	2	0	0	0	1	0	0	1	0	0	som
	Sulprofos S3	10; 1-1-1;1-1-12; 2-1-1;2-2-8;2-1-10;	1	0	0	0	1	0	0	1	1	0	2	1	0	0	nonsom
	Sulprofos S5	10; 1-1-12; 2-2-8;2-1-10;	1	0	0	0	0	0	0	1	0	0	2	1	0	0	nonsom
	Sulprofos S17	10; 1-1-1;1-1-3; 2-2-3;	1	0	0	0	1	0	1	0	0	2	0	0	0	0	som

**Fig 4. Converting the original Molprint2D text format into numerical values.** The original Molprint2D text format are converted to numerical values so that they are readable by the WEKA package. Such a converting for three selected training set compounds are shown in the figure. Label "I-" represents the neighbor type of the first layer while label "II-" represents the neighbor type of the second layer.

doi:10.1371/journal.pone.0169910.g004

In the quantum features,  $q_A(N+1)$ ,  $q_A(N-1)$ ,  $P_A(N)$ ,  $P_A(N+1)$  and  $P_A(N-1)$  were computed as condensed Fukui functions and were removed for further model training. Because some SYBYL atomic types were excluded from our dataset, the actual number of attributes used was 37 from layer 0 (central atom) to layer 2. Thus, the initial number of features created before feature selection were 42 (5 quantum mechanics features plus 37 attributes from circular fingerprints).

### Performance measurement

We used the grid search provided in the WEKA package to obtain the optimized parameters. The optimized SVM parameters  $C$ ,  $\gamma$  and  $W$  (weight) after the training process were 15, 1/ instances (default), and 1 (default), respectively. Normalization and the probability estimates were used and the other parameters employed were default settings. As shown in Table 3, the non-SOM set was designated as the class I while the SOM set was designated as the class II set. The performance of our training model was measured by sensitivity (SE), specificity (SP), accuracy (ACC), and Matthews correlation coefficient (MCC) computed which were 0.851, 0.864, 0.856, and 0.705, respectively (Table 3).

This would indicate that our model was well-balanced and could accurately differentiate the actual SOMs from the non-SOMs set since both SE and SP computed were high. The

**Table 3. The performance of the training and the test set given by the model built.**

Method		instances	Features	Parameters			SE	SP	ACC	MCC	AUC
				C 15	$\gamma$ 1/111	W 1					
SVM	Training set	111	5+37	C 15	$\gamma$ 1/111	W 1	0.851	0.864	0.856	0.705	0.887
	Test set	117	5+37	-	-	-	0.790	0.917	0.829	0.659	0.877
Naive Bayes	Training set	111	5+37	Default setting			0.821	0.864	0.838	0.673	0.869
	Test set	117	5+37	Default setting			0.790	0.889	0.821	0.635	0.867
Random Forest	Training set	111	5+37	Default setting			0.910	0.864	0.892	0.774	0.927
	Test set	117	5+37	Default setting			0.840	0.861	0.846	0.668	0.868

In the model training process, class I set was designated as the non-SOM while class II set was designated as the SOM set. Instances represent the number of potential SOMs identified. There were 5 quantum features and 37 attributes from circular fingerprints used for building the model. Parameters C,  $\gamma$  and W were obtained from the SVM training. The sensitivity (SE), specificity (SP), accuracy (ACC), Matthews correlation coefficient (MCC), and area under the ROC curve (AUC) were used to characterize the performance of the SVM model built. Symbol "-" was used to represent the same numbers used in the test set.

doi:10.1371/journal.pone.0169910.t003

Area Under Curve (AUC) of a ROC curve constructed could also reveal the effectiveness of a classification algorithm. A random model would give rise only a point along the diagonal line, while a good model would yield a point in the upper left corner of the ROC curve. Therefore, the AUC for a random and ideal model obtained would be 0.5 and 1.0, respectively. The ROC curve constructed for the test set using parameters obtained from the training set was presented in Fig 5. The AUC of the ROC curve constructed for the test set was 0.887, indicating again the high predicting accuracy of the model constructed. Some other classification methods such as Naive Bayes and Random Forest available from the WEKA [110] package were also employed to build the training models using the aforementioned



**Fig 5. The ROC curves of models.** The ROC curves constructed for (A) the training set by the SVM, (B) the test set by the SVM, (C) the training set by the Naive Bayes method, (D) the test set by the Naive Bayes method, (E) the training set by the Random Forest method, and (F) the test set by the Random Forest method.

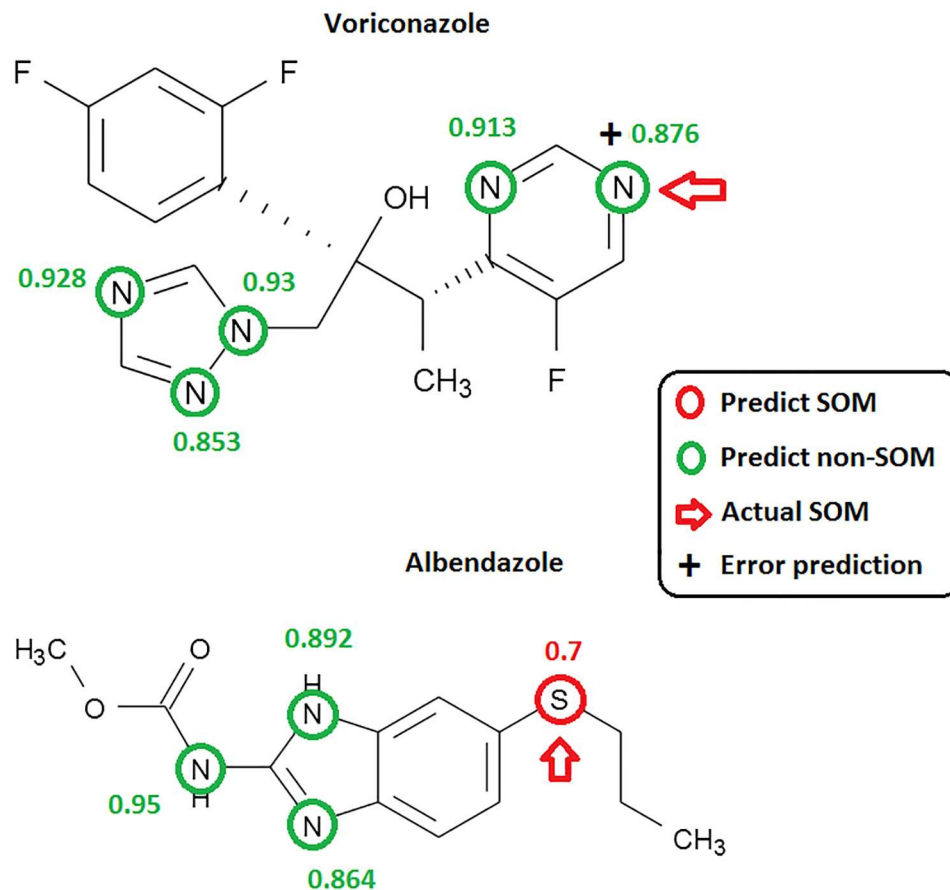
doi:10.1371/journal.pone.0169910.g005

features. The trained SVM model were also applied to the test set for constructing the ROC curves. As shown in Fig 5, the AUC under the ROC curves obtained from the training by Naive Bayes and Random Forest method with the default settings were respectively 0.869 and 0.927, which would reflect the number of features used was sufficient to represent the datasets. We also employed a test set to test the prediction accuracy of our SVM model constructed. This test set was composed of 117 instances including 40 SOMs and 65 non-SOMs which were collected from some substrates and non-substrates of FMO enzymes. The values of SE, SP, ACC, and MCC computed for the test set were 0.790, 0.917, 0.829, and 0.659, respectively (Table 3), since 33 out of 36 SOMs and 64 out of 81 non-SOMs were correctly predicted. As shown in Fig 5B, the AUC of the ROC curve computed for the test set was 0.877, indicating again both the features used and the SVM model constructed were adequate in predicting the SOMs of FMO enzymes. The prediction models built by both the Naive-Bayes and RandomForest methods were also cross-validated using the jackknife (leave-one-out) method. The corresponding AUC computed were 0.8911, 0.8667 and 0.9095 for SVM, NaiveBayes, and RandomForest, respectively (S1 Fig), indicating that all the three models built can accurately predict the SOMs.

As shown in Table 3, lower SE values in the test set obtained might be caused by bulky size or unfitted conformation of the ligands that were expelled from the enzyme active sites. For examples, no activity is observed for the phenothiazine derivative with an alkyl side chain of 3 carbon length (3PTZ in our data set), though  $K_m$  measured for carbons 5 and 8 derivatives (5PTZ and 8PTZ) were 18 and 10  $\mu\text{M}$ , respectively. The SVM model constructed was unable to lower the prediction error of non-SOM caused by these problems. The other factors which might affect the prediction accuracy of the test set would be the total number of instances used. At present, increasing the data size is still an impractical approach since there are not many human FMO substrates around and those identified for other species such as pig, rabbit, or rat are also scarce as well. The prediction probability of each potential SOM for two selected substrates voriconazole and albendazole was shown in Fig 6, while the total prediction probabilities for the test set were listed in S3 Table.

## Compare with the prediction results given by MetaPrint2D

MetaPrint2D [22] and Metasite [21] are the two known packages that can be also used to predict SOMs by Phase I enzymes. The former was a free online tool and was employed here on the same test set for comparing with our prediction results by the SVM models built. MetaPrint2D encodes each SOM and its corresponding substructure with circular fingerprints. Then, it screens a specific metabolic reaction and calculates the occurrence ratio (likelihood) of each SOM involved in the metabolic reaction. The number of exact levels and similarity threshold chosen were 3.0 and 0.5 which were both default settings. The weights of the six fingerprint levels used were 1.0, 1.0, 1.0, 0.75, 0.5, and 0.25, respectively. On output, each atom in a molecule was color coded by its normalized occurrence ratio (NOR) computed. Higher NOR computed means higher reported rate of being a metabolism site in the database. An atom was colored in red, orange, green, white, and grey to represent the corresponding NOR computed as 1.00~0.66, 0.66~0.33, 0.33~0.15, 0.15~0.00 and no data, respectively. Note that the substrates of all the phase I metabolic including FMOs enzymes were accounted in the training base by MetaPrint2D. We intended only to compare the true negative rate or SP (Eq 6) computed by MetaPrint2D with our SVM models since those predicted by the former would be much broader than those by the latter. Apparently, there were 22 out of 36 SOMs being correctly predicted by MetaPrint2D as shown in Table 4, If we defined atoms coded with red or orange color as being a potential SOM predicted by the



**Fig 6. The prediction probability of each potential SOM computed for two selected FMO substrates voriconazole and albendazole.** The actual SOMs of each compound determined are highlighted with red arrows. Each predicted SOM is marked with a red circle where the prediction probability computed is shown alongside. Each predicted non-SOM is marked by a green circle and the prediction probability computed is also shown alongside. Symbol "+" is used to denote a false prediction.

doi:10.1371/journal.pone.0169910.g006

method. However, the corresponding SP computed for the prediction result was 0.611 which was worse than that (0.917) given by our SVM model. This might be ascribed to the fact that fewer human FMO substrates was collected in MetaPrint2D than in our SVM model in the training procedure.

**Table 4. Comparison between the prediction results for the test set by the SVM model built and Metaprint2D.**

	Numbers of instances	Enzyme for SOM	TP	FN	SP
			FP	TN	
Test set	117	FMO	64	17	0.917
			3	33	
Metaprint2D	117	Phase I enzymes	63	18	0.611
			14	22	

Class I set is designated as the non-SOM while class II set is designated as the SOM set in the SVM model built. The number of substrates used in both sets is the same. The performance of the method is characterized by TP (True Positive), FP (False Positive), FN (False Negative), TN (True Negative), and specificity (SP) computed.

doi:10.1371/journal.pone.0169910.t004

## Conclusion

In this report, we have developed a SOM prediction method for some FMO enzymes using SVM with some quantum mechanics and circular fingerprints attributes. The total number of molecular descriptors used in building the SOM prediction model was 42, including 5 quantum mechanics features such as Mulliken charges, condensed FuKui Functions  $f_A^+$ ,  $f_A^-$ ,  $f_A^0$ , and vdw surface area and 37 circular fingerprints attributes. The prediction ability for SOM by the SVM model constructed was validated by both the training and the test sets and was found to be accurate. We have also compared the prediction result on the test set by our SVM model with that by a free online method MetaPrint2D. Unlike our SVM model which only focusing on FMO enzymes, all the phase I metabolic including FMO enzymes were considered by MetaPrint2D. There are some drugs that are actually not metabolized by CYP450 but rather by FMO enzymes such as benzydamine, itopride, and arbidol. Though FMO are also important phase I metabolic enzymes, it appears that far less attention has been paid to the enzymes than the CYP450 systems. Moreover, a major difference between FMO and CYP450 is that the adverse drug-drug interaction problem may be avoided for drugs that are solely metabolized by the former [7, 8]. Though the SVM prediction models constructed by us for only FMO enzymes may deserve merit, there are still rooms for improvement for our current models constructed especially in generating a correct bound conformation for a substrate. The current SVM methodology may be also extended for constructing the prediction models for the CYP450 substrates only.

## Supporting Information

**S1 Fig. The ROC curves of all the SVM, NaiveBayes, and RandomForest models built and cross-validated by the jackknife (leave-one-out) method are given.**

(TIFF)

**S1 Table. All the structures, actual SOMs and related references used in the dataset (the diagram of structures are obtained from open babel and NCBI) are listed in the table.**

(PDF)

**S2 Table. The original data set includes both the training and the test sets are given in the table.**

(XLS)

**S3 Table. The prediction probabilities computed for the test set are listed in the table.**

(XLSX)

## Acknowledgments

This work has been supported by a grant MOST 105-2320-B007-001-MY2 from the Ministry of Science and Technology, Taiwan ROC.

## Author Contributions

**Conceptualization:** CWF.

**Data curation:** CWF.

**Formal analysis:** CWF.

**Funding acquisition:** THL.

**Investigation:** CWF.

**Methodology:** CWF.

**Project administration:** THL.

**Resources:** THL.

**Software:** CWF.

**Supervision:** THL.

**Validation:** CWF THL.

**Visualization:** CWF.

**Writing – original draft:** CWF.

**Writing – review & editing:** THL.

## References

1. Dixit A, Roche TE. Spectrophotometric assay of the flavin-containing monooxygenase and changes in its activity in female mouse liver with nutritional and diurnal conditions. *Archives of biochemistry and biophysics*. 1984; 233(1):50–63. PMID: [6087744](#)
2. Cashman JR. Structural and Catalytic Properties of the Mammalian Flavin-Containing Monooxygenase. *Chem Res Toxicol*. 1995; 8(2):165–81.
3. Cashman JR. Human flavin-containing monooxygenase: substrate specificity and role in drug metabolism. *Current drug metabolism*. 2000; 1(2):181–91. PMID: [11465082](#)
4. Ziegler DM. Recent studies on the structure and function of multisubstrate flavin-containing monooxygenases. *Annual review of pharmacology and toxicology*. 1993; 33:179–99. doi: [10.1146/annurev.pa.33.040193.001143](#) PMID: [8494339](#)
5. Cashman JR. The implications of polymorphisms in mammalian flavin-containing monooxygenases in drug discovery and development. *Drug Discov Today*. 2004; 9(13):574–81. doi: [10.1016/S1359-6446\(04\)03136-8](#) PMID: [15203093](#)
6. Treacy EP, Akerman BR, Chow LML, Youil R, Bibeau C, Lin J, et al. Mutations of the flavin-containing monooxygenase gene (FMO3) cause trimethylaminuria, a defect in detoxication. *Hum Mol Genet*. 1998; 7(5):839–45. PMID: [9536088](#)
7. Cashman JR, Zhang J. Human flavin-containing monooxygenases. *Annual review of pharmacology and toxicology*. 2006; 46:65–100. doi: [10.1146/annurev.pharmtox.46.120604.141043](#) PMID: [16402899](#)
8. Phillips IR, Shephard EA. Flavin-containing monooxygenases: mutations, disease and drug response. *Trends in pharmacological sciences*. 2008; 29(6):294–301. doi: [10.1016/j.tips.2008.03.004](#) PMID: [18423897](#)
9. Capolongo F, Santi A, Anfossi P, Montesissa C. Benzydamine as a useful substrate of hepatic flavin-containing monooxygenase activity in veterinary species. *J Vet Pharmacol Ther*. 2010; 33(4):341–6. doi: [10.1111/j.1365-2885.2009.01145.x](#) PMID: [20646194](#)
10. Mushiroda T, Douya R, Takahara E, Nagata O. The involvement of flavin-containing monooxygenase but not CYP3A4 in metabolism of itopride hydrochloride, a gastroprokinetic agent: Comparison with cisapride and mosapride citrate. *Drug Metab Dispos*. 2000; 28(10):1231–7. PMID: [10997945](#)
11. Deng P, Zhong D, Yu K, Zhang Y, Wang T, Chen X. Pharmacokinetics, metabolism, and excretion of the antiviral drug arbidol in humans. *Antimicrobial agents and chemotherapy*. 2013; 57(4):1743–55. doi: [10.1128/AAC.02282-12](#) PMID: [23357765](#)
12. Ziegler DM. Flavin-containing monooxygenases: catalytic mechanism and substrate specificities. *Drug metabolism reviews*. 1988; 19(1):1–32. doi: [10.3109/03602538809049617](#) PMID: [3293953](#)
13. Krueger SK, Williams DE. Mammalian flavin-containing monooxygenases: structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacology & therapeutics*. 2005; 106(3):357–87.
14. Goracci L, Ceccarelli M, Bonelli D, Cruciani G. Modeling phospholipidosis induction: reliability and warnings. *Journal of chemical information and modeling*. 2013; 53(6):1436–46. doi: [10.1021/ci400113t](#) PMID: [23692521](#)

15. Kousba A, Soll R, Yee S, Martin M. Cyclic conversion of the novel src kinase inhibitor [7-(2,6Dichlorophenyl)-5-methyl-benzo[1,2,4]triazin-3-yl]-[4-(2-pyrrolidin-1-yl-ethoxy)-phenyl]-amine(TG100435) and its n-oxide metabolite by flavin-containing monooxygenases and cytochrome p450 reductase. *Drug Metab Dispos.* 2007; 35(12):2242–51. doi: [10.1124/dmd.107.017384](https://doi.org/10.1124/dmd.107.017384) PMID: [17881660](https://pubmed.ncbi.nlm.nih.gov/17881660/)
16. Fura A. Role of pharmacologically active metabolites in drug discovery and development. *Drug Discov Today.* 2006; 11(3–4):133–42. doi: [10.1016/S1359-6446\(05\)03681-0](https://doi.org/10.1016/S1359-6446(05)03681-0) PMID: [16533711](https://pubmed.ncbi.nlm.nih.gov/16533711/)
17. Rydberg P, Gloriam DE, Zaretski J, Breneman C, Olsen L. SMARTCyp: A 2D Method for Prediction of Cytochrome P450-Mediated Drug Metabolism. *ACS medicinal chemistry letters.* 2010; 1(3):96–100. doi: [10.1021/ml100016x](https://doi.org/10.1021/ml100016x) PMID: [24936230](https://pubmed.ncbi.nlm.nih.gov/24936230/)
18. Campagna-Slater V, Pottel J, Therrien E, Cantin LD, Moitessier N. Development of a Computational Tool to Rival Experts in the Prediction of Sites of Metabolism of Xenobiotics by P450s. *Journal of chemical information and modeling.* 2012; 52(9):2471–83. doi: [10.1021/ci3003073](https://doi.org/10.1021/ci3003073) PMID: [22916680](https://pubmed.ncbi.nlm.nih.gov/22916680/)
19. Tyzack JD, Williamson MJ, Torella R, Glen RC. Prediction of cytochrome P450 xenobiotic metabolism: tethered docking and reactivity derived from ligand molecular orbital analysis. *Journal of chemical information and modeling.* 2013; 53(6):1294–305. doi: [10.1021/ci400058s](https://doi.org/10.1021/ci400058s) PMID: [23701380](https://pubmed.ncbi.nlm.nih.gov/23701380/)
20. Jones G, Willett P, Glen RC. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *Journal of molecular biology.* 1995; 245(1):43–53. PMID: [7823319](https://pubmed.ncbi.nlm.nih.gov/7823319/)
21. Cruciani G, Valeri A, Goracci L, Pellegrino RM, Buonerba F, Baroni M. Flavin monooxygenase metabolism: why medicinal chemists should matter. *Journal of medicinal chemistry.* 2014; 57(14):6183–96. doi: [10.1021/jm5007098](https://doi.org/10.1021/jm5007098) PMID: [25003501](https://pubmed.ncbi.nlm.nih.gov/25003501/)
22. Carlsson L, Spjuth O, Adams S, Glen RC, Boyer S. Use of historic metabolic biotransformation data as a means of anticipating metabolic sites using MetaPrint2D and Bioclipse. *BMC bioinformatics.* 2010; 11:362. doi: [10.1186/1471-2105-11-362](https://doi.org/10.1186/1471-2105-11-362) PMID: [20594327](https://pubmed.ncbi.nlm.nih.gov/20594327/)
23. Accelrys Inc. TC, Suite 100, San Diego, CA, 92121, USA. Accelrys Metabolite Database. [<http://accelrys.com/products/databases/bioactivity/metabolite.html>].
24. Zaretski J, Bergeron C, Rydberg P, Huang TW, Bennett KP, Breneman CM. RS-predictor: a new tool for predicting sites of cytochrome P450-mediated metabolism applied to CYP 3A4. *Journal of chemical information and modeling.* 2011; 51(7):1667–89. doi: [10.1021/ci2000488](https://doi.org/10.1021/ci2000488) PMID: [21528931](https://pubmed.ncbi.nlm.nih.gov/21528931/)
25. Zaretski J, Rydberg P, Bergeron C, Bennett KP, Olsen L, Breneman CM. RS-Predictor models augmented with SMARTCyp reactivities: robust metabolic regioselectivity predictions for nine CYP isozymes. *Journal of chemical information and modeling.* 2012; 52(6):1637–59. doi: [10.1021/ci300009z](https://doi.org/10.1021/ci300009z) PMID: [22524152](https://pubmed.ncbi.nlm.nih.gov/22524152/)
26. Zaretski J, Matlock M, Swamidass SJ. XenoSite: accurately predicting CYP-mediated sites of metabolism with neural networks. *Journal of chemical information and modeling.* 2013; 53(12):3373–83. doi: [10.1021/ci400518g](https://doi.org/10.1021/ci400518g) PMID: [24224933](https://pubmed.ncbi.nlm.nih.gov/24224933/)
27. Daylight Chemical Information Systems IAV, CA. [<http://www.daylight.com/dayhtml/doc/theory/theory.finger.html>].
28. Mu F, Unkefer CJ, Unkefer PJ, Hlavacek WS. Prediction of metabolic reactions based on atomic and molecular properties of small-molecule compounds. *Bioinformatics.* 2011; 27(11):1537–45. doi: [10.1093/bioinformatics/btr177](https://doi.org/10.1093/bioinformatics/btr177) PMID: [21478194](https://pubmed.ncbi.nlm.nih.gov/21478194/)
29. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research.* 2012; 40(Database issue):D109–14. doi: [10.1093/nar/gkr988](https://doi.org/10.1093/nar/gkr988) PMID: [22080510](https://pubmed.ncbi.nlm.nih.gov/22080510/)
30. Kirchmair J, Williamson MJ, Afzal AM, Tyzack JD, Choy AP, Howlett A, et al. FAME: A rapid and accurate predictor of sites of metabolism in multiple species by endogenous enzymes. *Journal of chemical information and modeling.* 2013; 53(11):2896–907. doi: [10.1021/ci400503s](https://doi.org/10.1021/ci400503s) PMID: [24219364](https://pubmed.ncbi.nlm.nih.gov/24219364/)
31. Steinbeck C, Hoppe C, Kuhn S, Floris M, Guha R, Willighagen EL. Recent developments of the chemistry development kit (CDK)—an open-source java library for chemo- and bioinformatics. *Current pharmaceutical design.* 2006; 12(17):2111–20. PMID: [16796559](https://pubmed.ncbi.nlm.nih.gov/16796559/)
32. Olsen L, Rydberg P, Rod TH, Ryde U. Prediction of activation energies for hydrogen abstraction by cytochrome p450. *Journal of medicinal chemistry.* 2006; 49(22):6489–99. doi: [10.1021/jm060551l](https://doi.org/10.1021/jm060551l) PMID: [17064067](https://pubmed.ncbi.nlm.nih.gov/17064067/)
33. Rydberg P, Ryde U, Olsen L. Prediction of activation energies for aromatic oxidation by cytochrome P450. *The journal of physical chemistry A.* 2008; 112(50):13058–65. doi: [10.1021/jp803854v](https://doi.org/10.1021/jp803854v) PMID: [18986131](https://pubmed.ncbi.nlm.nih.gov/18986131/)
34. Molecular Operating Environment (MOE) CCGI, 1010 Sherbooke St. West, Suite 910, Montreal, QC, Canada, H3A 2R7, 2012. [[https://www.chemcomp.com/MOEMolecular\\_Operating\\_Environment.htm](https://www.chemcomp.com/MOEMolecular_Operating_Environment.htm)].



35. Hasegawa K, Koyama M, Funatsu K. Quantitative Prediction of Regioselectivity Toward Cytochrome P450/3A4 Using Machine Learning Approaches. *Mol Inform.* 2010; 29(3):243–9. doi: [10.1002/minf.200900086](https://doi.org/10.1002/minf.200900086) PMID: [27462767](https://pubmed.ncbi.nlm.nih.gov/27462767/)
36. Rudik AV, Dmitriev AV, Lagunin AA, Filimonov DA, Poroikov VV. Metabolism site prediction based on xenobiotic structural formulas and PASS prediction algorithm. *Journal of chemical information and modeling.* 2014; 54(2):498–507. doi: [10.1021/ci400472j](https://doi.org/10.1021/ci400472j) PMID: [24417355](https://pubmed.ncbi.nlm.nih.gov/24417355/)
37. Tyzack JD, Mussa HY, Williamson MJ, Kirchmair J, Glen RC. Cytochrome P450 site of metabolism prediction from 2D topological fingerprints using GPU accelerated probabilistic classifiers. *Journal of cheminformatics.* 2014; 6:29. doi: [10.1186/1758-2946-6-29](https://doi.org/10.1186/1758-2946-6-29) PMID: [24959208](https://pubmed.ncbi.nlm.nih.gov/24959208/)
38. Xing L, Glen RC. Novel methods for the prediction of logP, pK(a), and logD. *Journal of chemical information and computer sciences.* 2002; 42(4):796–805. PMID: [12132880](https://pubmed.ncbi.nlm.nih.gov/12132880/)
39. Xing L, Glen RC, Clark RD. Predicting pK(a) by molecular tree structured fingerprints and PLS. *Journal of chemical information and computer sciences.* 2003; 43(3):870–9. doi: [10.1021/ci020386s](https://doi.org/10.1021/ci020386s) PMID: [12767145](https://pubmed.ncbi.nlm.nih.gov/12767145/)
40. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, et al. HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic acids research.* 2013; 41(Database issue):D801–7. doi: [10.1093/nar/gks1065](https://doi.org/10.1093/nar/gks1065) PMID: [23161693](https://pubmed.ncbi.nlm.nih.gov/23161693/)
41. Ripp SL, Itagaki K, Philpot RM, Elfarra AA. Methionine S-oxidation in human and rabbit liver microsomes: evidence for a high-affinity methionine S-oxidase activity that is distinct from flavin-containing monooxygenase 3. *Archives of biochemistry and biophysics.* 1999; 367(2):322–32. doi: [10.1006/abbi.1999.1247](https://doi.org/10.1006/abbi.1999.1247) PMID: [10395751](https://pubmed.ncbi.nlm.nih.gov/10395751/)
42. Lang DH, Yeung CK, Peter RM, Ibarra C, Gasser R, Itagaki K, et al. Isoform specificity of trimethylamine N-oxygenation by human flavin-containing monooxygenase (FMO) and P450 enzymes: selective catalysis by FMO3. *Biochemical pharmacology.* 1998; 56(8):1005–12. PMID: [9776311](https://pubmed.ncbi.nlm.nih.gov/9776311/)
43. Cashman JR, Xiong YN, Xu L, Janowsky A. N-oxygenation of amphetamine and methamphetamine by the human flavin-containing monooxygenase (form 3): role in bioactivation and detoxication. *The Journal of pharmacology and experimental therapeutics.* 1999; 288(3):1251–60. PMID: [10027866](https://pubmed.ncbi.nlm.nih.gov/10027866/)
44. Lin J, Berkman CE, Cashman JR. N-oxygenation of primary amines and hydroxylamines and retroreduction of hydroxylamines by adult human liver microsomes and adult human flavin-containing monooxygenase 3. *Chem Res Toxicol.* 1996; 9(7):1183–93. doi: [10.1021/tx9600614](https://doi.org/10.1021/tx9600614) PMID: [8902275](https://pubmed.ncbi.nlm.nih.gov/8902275/)
45. Lin J, Cashman JR. N-oxygenation of phenethylamine to the trans-oxime by adult human liver flavin-containing monooxygenase and retroreduction of phenethylamine hydroxylamine by human liver microsomes. *The Journal of pharmacology and experimental therapeutics.* 1997; 282(3):1269–79. PMID: [9316835](https://pubmed.ncbi.nlm.nih.gov/9316835/)
46. Lin J, Cashman JR. Detoxication of tyramine by the flavin-containing monooxygenase: stereoselective formation of the trans oxime. *Chem Res Toxicol.* 1997; 10(8):842–52. doi: [10.1021/tx970030o](https://doi.org/10.1021/tx970030o) PMID: [9282832](https://pubmed.ncbi.nlm.nih.gov/9282832/)
47. Rodriguez RJ, Miranda CL. Isoform specificity of N-deacetyl ketoconazole by human and rabbit flavin-containing monooxygenases. *Drug Metab Dispos.* 2000; 28(9):1083–6. PMID: [10950853](https://pubmed.ncbi.nlm.nih.gov/10950853/)
48. Rodrigues AD, Kukulka MJ, Ferrero JL, Cashman JR. In vitro hepatic metabolism of ABT-418 in chimpanzee (*Pan troglodytes*). A unique pattern of microsomal flavin-containing monooxygenase-dependent stereoselective N'-oxidation. *Drug Metab Dispos.* 1995; 23(10):1143–52. PMID: [8654204](https://pubmed.ncbi.nlm.nih.gov/8654204/)
49. Lang DH, Rettie AE. In vitro evaluation of potential in vivo probes for human flavin-containing monooxygenase (FMO): metabolism of benzydamine and caffeine by FMO and P450 isoforms. *British journal of clinical pharmacology.* 2000; 50(4):311–4. doi: [10.1046/j.1365-2125.2000.00265.x](https://doi.org/10.1046/j.1365-2125.2000.00265.x) PMID: [11012553](https://pubmed.ncbi.nlm.nih.gov/11012553/)
50. Tugnait M, Hawes EM, McKay G, Rettie AE, Haining RL, Midha KK. N-oxygenation of clozapine by flavin-containing monooxygenase. *Drug Metab Dispos.* 1997; 25(4):524–7. PMID: [9107553](https://pubmed.ncbi.nlm.nih.gov/9107553/)
51. Mushiroda T, Douya R, Takahara E, Nagata O. The involvement of flavin-containing monooxygenase but not CYP3A4 in metabolism of itopride hydrochloride, a gastroprokinetic agent: comparison with cisapride and mosapride citrate. *Drug Metab Dispos.* 2000; 28(10):1231–7. PMID: [10997945](https://pubmed.ncbi.nlm.nih.gov/10997945/)
52. Jacobsen W, Christians U, Benet LZ. In vitro evaluation of the disposition of A novel cysteine protease inhibitor. *Drug Metab Dispos.* 2000; 28(11):1343–51. PMID: [11038163](https://pubmed.ncbi.nlm.nih.gov/11038163/)
53. Park SB, Jacob P 3rd, Benowitz NL, Cashman JR. Stereoselective metabolism of (S)-(-)-nicotine in humans: formation of trans-(S)-(-)-nicotine N-1'-oxide. *Chem Res Toxicol.* 1993; 6(6):880–8. PMID: [8117928](https://pubmed.ncbi.nlm.nih.gov/8117928/)
54. Kajita J, Inano K, Fuse E, Kuwabara T, Kobayashi H. Effects of olopatadine, a new antiallergic agent, on human liver microsomal cytochrome P450 activities. *Drug Metab Dispos.* 2002; 30(12):1504–11. PMID: [12433826](https://pubmed.ncbi.nlm.nih.gov/12433826/)

55. Washio T, Arisawa H, Kohsaka K, Yasuda H. Identification of human drug-metabolizing enzymes involved in the metabolism of SNI-2011. *Biological & pharmaceutical bulletin*. 2001; 24(11):1263–6.
56. Ring BJ, Wrighton SA, Aldridge SL, Hansen K, Haehner B, Shipley LA. Flavin-containing monooxygenase-mediated N-oxidation of the M(1)-muscarinic agonist xanomeline. *Drug Metab Dispos*. 1999; 27(10):1099–103. PMID: [10497134](#)
57. Henderson MC, Krueger SK, Siddens LK, Stevens JF, Williams DE. S-oxygenation of the thioether organophosphate insecticides phorate and disulfoton by human lung flavin-containing monooxygenase 2. *Biochemical pharmacology*. 2004; 68(5):959–67. doi: [10.1016/j.bcp.2004.05.051](#) PMID: [15294458](#)
58. Furnes B, Schlenk D. Extrahepatic metabolism of carbamate and organophosphate thioether compounds by the flavin-containing monooxygenase and cytochrome P450 systems. *Drug Metab Dispos*. 2005; 33(2):214–8. doi: [10.1124/dmd.104.000984](#) PMID: [15547051](#)
59. Pike MG, Mays DC, Macomber DW, Lipsky JJ. Metabolism of a disulfiram metabolite, S-methyl N,N-diethylthiocarbamate, by flavin monooxygenase in human renal microsomes. *Drug Metab Dispos*. 2001; 29(2):127–32. PMID: [11159801](#)
60. Ohmi N, Yoshida H, Endo H, Hasegawa M, Akimoto M, Higuchi S. S-oxidation of S-methyl-esonari-mod by flavin-containing monooxygenases in human liver microsomes. *Xenobiotica; the fate of foreign compounds in biological systems*. 2003; 33(12):1221–31. doi: [10.1080/00498250310001624627](#) PMID: [14742144](#)
61. Karanam BV, Welch CJ, Reddy VG, Chilenski J, Biba M, Vincent S. Species differential stereoselective oxidation of a methylsulfide metabolite of MK-0767 [(+/-)-5-[(2,4-dioxothiazolidin-5-yl)methyl]-2-methoxy-N-[[4-(trifluoromethyl)phenyl]methyl]benzamide], a peroxisome proliferator-activated receptor dual agonist. *Drug Metab Dispos*. 2004; 32(10):1061–8. doi: [10.1124/dmd.104.000224](#) PMID: [15229170](#)
62. Hamman MA, Haehner-Daniels BD, Wrighton SA, Rettie AE, Hall SD. Stereoselective sulfoxidation of sulindac sulfide by flavin-containing monooxygenases. Comparison of human liver and kidney microsomes and mammalian enzymes. *Biochemical pharmacology*. 2000; 60(1):7–17. PMID: [10807940](#)
63. Attar M, Dong D, Ling KH, Tang-Liu DD. Cytochrome P450 2C8 and flavin-containing monooxygenases are involved in the metabolism of tazarotenic acid in humans. *Drug Metab Dispos*. 2003; 31(4):476–81. PMID: [12642475](#)
64. Yeung CK, Lang DH, Thummel KE, Rettie AE. Immunoquantitation of FMO1 in human liver, kidney, and intestine. *Drug Metab Dispos*. 2000; 28(9):1107–11. PMID: [10950857](#)
65. Krueger SK, Martin SR, Yueh MF, Pereira CB, Williams DE. Identification of active flavin-containing monooxygenase isoform 2 in human lung and characterization of expressed protein. *Drug Metab Dispos*. 2002; 30(1):34–41. PMID: [11744609](#)
66. Henderson MC, Krueger SK, Stevens JF, Williams DE. Human flavin-containing monooxygenase form 2 S-oxygenation: sulfenic acid formation from thioureas and oxidation of glutathione. *Chem Res Toxicol*. 2004; 17(5):633–40. doi: [10.1021/tx034253s](#) PMID: [15144220](#)
67. Kim YM, Ziegler DM. Size limits of thiocarbamides accepted as substrates by human flavin-containing monooxygenase 1. *Drug Metab Dispos*. 2000; 28(8):1003–6. PMID: [10901713](#)
68. Zhou S, Kestell P, Paxton JW. 6-methylhydroxylation of the anti-cancer agent 5,6-dimethylxanthone-4-acetic acid (DMXAA) by flavin-containing monooxygenase 3. *European journal of drug metabolism and pharmacokinetics*. 2002; 27(3):179–83. PMID: [12365199](#)
69. Chung WG, Cha YN. Oxidation of caffeine to theobromine and theophylline is catalyzed primarily by flavin-containing monooxygenase in liver microsomes. *Biochemical and biophysical research communications*. 1997; 235(3):685–8. doi: [10.1006/bbrc.1997.6866](#) PMID: [9207220](#)
70. Frederick CB, Mays JB, Ziegler DM, Guengerich FP, Kadlubar FF. Cytochrome P-450- and flavin-containing monooxygenase-catalyzed formation of the carcinogen N-hydroxy-2-aminofluorene and its covalent binding to nuclear DNA. *Cancer research*. 1982; 42(7):2671–7. PMID: [7083159](#)
71. Chung WG, Park CS, Roh HK, Lee WK, Cha YN. Oxidation of ranitidine by isozymes of flavin-containing monooxygenase and cytochrome P450. *Japanese journal of pharmacology*. 2000; 84(2):213–20. PMID: [11128045](#)
72. Shiraga T, Yajima K, Teragaki T, Suzuki K, Hashimoto T, Iwatsubo T, et al. Identification of enzymes responsible for the N-oxidation of darexaban glucuronide, the pharmacologically active metabolite of darexaban, and the glucuronidation of darexaban N-oxides in human liver microsomes. *Biological & pharmaceutical bulletin*. 2012; 35(3):413–21.
73. Hanlon SP, Camattari A, Abad S, Glieder A, Kittelmann M, Lutz S, et al. Expression of recombinant human flavin monooxygenase and moclobemide-N-oxide synthesis on multi-mg scale. *Chemical communications*. 2012.

74. Cashman JR, Ziegler DM. Contribution of N-oxygenation to the metabolism of MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) by various liver preparations. *Molecular pharmacology*. 1986; 29(2):163–7. PMID: [3485239](#)
75. Ripp SL, Overby LH, Philpot RM, Elfarra AA. Oxidation of cysteine S-conjugates by rabbit liver microsomes and cDNA-expressed flavin-containing mono-oxygenases: studies with S-(1,2-dichlorovinyl)-L-cysteine, S-(1,2,2-trichlorovinyl)-L-cysteine, S-allyl-L-cysteine, and S-benzyl-L-cysteine. *Molecular pharmacology*. 1997; 51(3):507–15. PMID: [9058607](#)
76. Yamazaki M, Shimizu M, Uno Y, Yamazaki H. Drug oxygenation activities mediated by liver microsomal flavin-containing monooxygenases 1 and 3 in humans, monkeys, rats, and minipigs. *Biochemical pharmacology*. 2014; 90(2):159–65. doi: [10.1016/j.bcp.2014.04.019](#) PMID: [24821112](#)
77. Catucci G, Occhipinti A, Maffei M, Gilardi G, Sadeghi SJ. Effect of human flavin-containing monooxygenase 3 polymorphism on the metabolism of aurora kinase inhibitors. *International journal of molecular sciences*. 2013; 14(2):2707–16. doi: [10.3390/ijms14022707](#) PMID: [23358255](#)
78. Henderson MC, Siddens LK, Krueger SK, Stevens JF, Kedzie K, Fang WK, et al. Flavin-containing monooxygenase S-oxygenation of a series of thioureas and thiones. *Toxicology and applied pharmacology*. 2014; 278(2):91–9. doi: [10.1016/j.taap.2014.04.002](#) PMID: [24727368](#)
79. Pichard-Garcia L, Weaver RJ, Eckett N, Scarfe G, Fabre JM, Lucas C, et al. The olivacine derivative 16020 (9-hydroxy-5,6-dimethyl-N-[2-(dimethylamino)ethyl]-6H-pyrido(4,3-B)-carbazole-1-carboxamide) induces CYP1A and its own metabolism in human hepatocytes in primary culture. *Drug Metab Dispos*. 2004; 32(1):80–8. doi: [10.1124/dmd.32.1.80](#) PMID: [14709624](#)
80. Usmani KA, Karoly ED, Hodgson E, Rose RL. In vitro sulfoxidation of thioether compounds by human cytochrome P450 and flavin-containing monooxygenase isoforms with particular reference to the CYP2C subfamily. *Drug Metab Dispos*. 2004; 32(3):333–9. doi: [10.1124/dmd.32.3.333](#) PMID: [14977868](#)
81. Vyas PM, Roychowdhury S, Koukouritaki SB, Hines RN, Krueger SK, Williams DE, et al. Enzyme-mediated protein haptentation of dapsone and sulfamethoxazole in human keratinocytes: II. Expression and role of flavin-containing monooxygenases and peroxidases. *The Journal of pharmacology and experimental therapeutics*. 2006; 319(1):497–505. doi: [10.1124/jpet.106.105874](#) PMID: [16857727](#)
82. Yanni SB, Annaert PP, Augustijns P, Ibrahim JG, Benjamin DK Jr., Thakker DR. In vitro hepatic metabolism explains higher clearance of voriconazole in children versus adults: role of CYP2C19 and flavin-containing monooxygenase 3. *Drug Metab Dispos*. 2010; 38(1):25–31. doi: [10.1124/dmd.109.029769](#) PMID: [19841059](#)
83. Ring BJ, Catlow J, Lindsay TJ, Gillespie T, Roskos LK, Cerimele BJ, et al. Identification of the human cytochromes P450 responsible for the in vitro formation of the major oxidative metabolites of the antipsychotic agent olanzapine. *The Journal of pharmacology and experimental therapeutics*. 1996; 276(2):658–66. PMID: [8632334](#)
84. Lee SK, Kang MJ, Jin C, In MK, Kim DH, Yoo HH. Flavin-containing monooxygenase 1-catalysed N, N-dimethylamphetamine N-oxidation. *Xenobiotica; the fate of foreign compounds in biological systems*. 2009; 39(9):680–6. doi: [10.1080/00498250902998699](#) PMID: [19552509](#)
85. Salva M, Jansat JM, Martinez-Tobed A, Palacios JM. Identification of the human liver enzymes involved in the metabolism of the antimigraine agent almotriptan. *Drug Metab Dispos*. 2003; 31(4):404–11. PMID: [12642466](#)
86. Wang L, Christopher LJ, Cui D, Li W, Iyer R, Humphreys WG, et al. Identification of the human enzymes involved in the oxidative metabolism of dasatinib: an effective approach for determining metabolite formation kinetics. *Drug Metab Dispos*. 2008; 36(9):1828–39. doi: [10.1124/dmd.107.020255](#) PMID: [18556438](#)
87. Gazquez JL, Mendez F. The Hard and Soft Acids and Bases Principle—An Atoms in Molecules Viewpoint. *J Phys Chem-U.S.* 1994; 98(17):4591–3.
88. Mendez F, Gazquez JL. Chemical-Reactivity of Enolate Ions—the Local Hard and Soft Acids and Bases Principle Viewpoint. *J Am Chem Soc*. 1994; 116(20):9298–301.
89. Parr RG, Yang WT. Density Functional-Approach to the Frontier-Electron Theory of Chemical-Reactivity. *J Am Chem Soc*. 1984; 106(14):4049–50.
90. Yang W, Parr RG, Pucci R. Electron-Density, Kohn-Sham Frontier Orbitals, and Fukui Functions. *J Chem Phys*. 1984; 81(6):2862–3.
91. Ayers PW, Levy M. Perspective on "Density functional approach to the frontier-electron theory of chemical reactivity"—Parr RG, Yang W (1984) *J Am Chem Soc* 106: 4049–4050. *Theor Chem Acc*. 2000; 103(3–4):353–60.
92. Ayers P.W. Y W, Bartolotti L.J., In: Chattaraj P.K. (Ed.). *Chemical reactivity theory: A density functional view*, CRC Press, Boca Raton, 2009.

93. Fukui K, Yonezawa T, Shingu H. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *J Chem Phys.* 1952; 20(4):722–5.
94. Yang W, Mortier WJ. The Use of Global and Local Molecular-Parameters for the Analysis of the Gas-Phase Basicity of Amines. *J Am Chem Soc.* 1986; 108(19):5708–11. doi: [10.1021/ja00279a008](https://doi.org/10.1021/ja00279a008) PMID: [22175316](https://pubmed.ncbi.nlm.nih.gov/22175316/)
95. Tiznado W, Chamorro E, Contreras R, Fuentealba P. Comparison among four different ways to condense the Fukui function. *Journal of Physical Chemistry A.* 2005; 109(14):3220–4.
96. Ayers PW, Morrison RC, Roy RK. Variational principles for describing chemical reactions: Condensed reactivity indices. *J Chem Phys.* 2002; 116(20):8731–44.
97. Bultinck P, Fias S, Van Alsenoy C, Ayers PW, Carbo-Dorca R. Critical thoughts on computing atom condensed Fukui functions. *J Chem Phys.* 2007; 127(3).
98. Bultinck P, Langenaeker W, Carbo-Dorca R, Tollenaere JP. Fast calculation of quantum chemical molecular descriptors from the electronegativity equalization method. *Journal of chemical information and computer sciences.* 2003; 43(2):422–8. doi: [10.1021/ci0255883](https://doi.org/10.1021/ci0255883) PMID: [12653504](https://pubmed.ncbi.nlm.nih.gov/12653504/)
99. Becke AD. Density-Functional Thermochemistry .3. The Role of Exact Exchange. *J Chem Phys.* 1993; 98(7):5648–52.
100. Becke AD. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys Rev A.* 1988; 38(6):3098–100.
101. Lee CT, Yang WT, Parr RG. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys Rev B.* 1988; 37(2):785–9.
102. Krishnan R, Binkley JS, Seeger R, Pople JA. Self-Consistent Molecular-Orbital Methods .20. Basis Set for Correlated Wave-Functions. *J Chem Phys.* 1980; 72(1):650–4.
103. Mclean AD, Chandler GS. Contracted Gaussian-Basis Sets for Molecular Calculations .1. 2nd Row Atoms, Z = 11–18. *J Chem Phys.* 1980; 72(10):5639–48.
104. Frisch M.J. TGW, Schlegel H.B., Scuseria G.E., Robb, M.A. CJRea. Gaussian 09, Revision C.01., Gaussian, Inc., Wallingford CT, 2009.
105. Bender A, Mussa HY, Glen RC, Reiling S. Molecular similarity searching using atom environments, information-based feature selection, and a naive Bayesian classifier. *Journal of chemical information and computer sciences.* 2004; 44(1):170–8. doi: [10.1021/ci034207y](https://doi.org/10.1021/ci034207y) PMID: [14741025](https://pubmed.ncbi.nlm.nih.gov/14741025/)
106. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *Journal of cheminformatics.* 2011; 3.
107. Burges CJC. A tutorial on Support Vector Machines for pattern recognition. *Data Min Knowl Disc.* 1998; 2(2):121–67.
108. Vapnik VN. An overview of statistical learning theory. *IEEE T Neural Networ.* 1999; 10(5):988–99.
109. Chang CC, Lin CJ. LIBSVM: A Library for Support Vector Machines. *Acm T Intel Syst Tec.* 2011; 2(3).
110. Hall Mark F E, Holmes Geoffrey, Pfahringer Bernhard, Reutemann Peter, H Ian. The WEKA Data Mining Software: An Update. *SIGKDD Explorations.* 2009; 11(1).
111. Chou KC, Zhang CT. Prediction of protein structural classes. *Critical reviews in biochemistry and molecular biology.* 1995; 30(4):275–349. doi: [10.3109/10409239509083488](https://doi.org/10.3109/10409239509083488) PMID: [7587280](https://pubmed.ncbi.nlm.nih.gov/7587280/)
112. Lin H, Ding H, Guo FB, Zhang AY, Huang J. Predicting subcellular localization of mycobacterial proteins by using Chou's pseudo amino acid composition. *Protein and peptide letters.* 2008; 15(7):739–44. PMID: [18782071](https://pubmed.ncbi.nlm.nih.gov/18782071/)
113. Tang H, Zou P, Zhang C, Chen R, Chen W, Lin H. Identification of apolipoprotein using feature selection technique. *Scientific reports.* 2016; 6:30441. doi: [10.1038/srep30441](https://doi.org/10.1038/srep30441) PMID: [27443605](https://pubmed.ncbi.nlm.nih.gov/27443605/)
114. Liu B, Fang L, Wang S, Wang X, Li H, Chou KC. Identification of microRNA precursor with the degenerate K-tuple or Kmer strategy. *Journal of theoretical biology.* 2015; 385:153–9. doi: [10.1016/j.jtbi.2015.08.025](https://doi.org/10.1016/j.jtbi.2015.08.025) PMID: [26362104](https://pubmed.ncbi.nlm.nih.gov/26362104/)