Quantitative Structure-Activity Relationship Modeling to Predict Drug-Drug Interactions Between Acetaminophen and Ingredients in Energy Drinks

by

Emese Somogyvari

A thesis submitted to the
School of Computing
in conformity with the requirements for
the degree of Master of Science

Queen's University
Kingston, Ontario, Canada
August 2014

Abstract

The evaluation of drug-drug interactions (DDI) is a crucial step in pharmaceutical drug discovery and design. Unfortunately, if adverse effects are to occur between the co-administration of two or more drugs, they are often difficult to test for. Traditional methods rely on *in vitro* studies as a basis for further *in vivo* assessment which can be a slow and costly process that may not detect all interactions. Here is presented a quantitative structure-activity relationship (QSAR) modeling approach that may be used to screen drugs early in development and bring new, beneficial drugs to market more quickly and at a lesser cost. A data set of 6,532 drugs was obtained from DrugBank for which 292 QSAR descriptors were calculated. The multi-label support vector machines (SVM) method was used for classification and the K-means method was used to cluster the data. The model was validated *in vitro* by exposing Hepa1-6 cells to select compounds found in energy drinks and assessing cell death. Model accuracy was found to be 99%, predicting 50% of known interactions despite being biased to predicting non-interacting drug pairs. Cluster analysis revealed interesting information, although current progress shows that more data is needed to better analyse results, and tools that bring various drug information together would be beneficial. Non-transfected Hepa1-6 cells exposed to acetaminophen, pyridoxine, creatine, L-carnitine, taurine and caffeine did not reveal any significant drug-drug

interactions, nor were they predicted by the model.

Co-Authorship

This research was conducted by Emese Somogyvari, under the supervision of Dr. Selim G. Akl and Dr. Louise M. Winn.

Acknowledgments

I would like to thank my supervisors Dr. Selim Akl of the School of Computing, Queens University and Dr. Louise Winn of the Department of Biomedical and Molecular Sciences, Queens University for their support and guidance throughout this project, as well as faithful lab members Chrissy, Christina, Margo, Nikki, and Sam for their patience and constant assistance.

Contents

List of Tables

List of Figures

List of Abbreviations

β-NADH β-Nicotinamide Adenine Dinucleotide

ADMET Absorption, distribution, metabolism, excretion and transport

ADR Adverse Drug Reaction

AERS Adverse Event Reporting System

ANOVA Analysis of Variance

ATCC American Type Culture Collection

BSA Bovine Serum Albumin

CDK Chemistry Development Kit

DDI Drug-Drug Interaction

DMEM Dulbecco's Modified Eagle's Serum

DT Decision Trees

FBS Fetal Bovine Serum

FDA Food and Drug Administration

IPF Interaction Profile Fingerprint

kNN Kappa Nearest Neighbour

LDH Lactate Dehydrogenase

NAPQI N-acetyl-p-benzoquinone-imine

NCE New Chemical Entity

OvR One-Vs.Rest

P450 Cytochrome P450

PB Phosphate Buffer

PBS Phosphate Buffered Saline

PCA Principal Components Analysis

QSAR Quantitative Structure-Activity Relationship

RBF Radial Basis Function

RF Random Forest

SDS-PAGE Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis

SRS Spontaneous Reporting System

SVM Support Vector Machines

TBS-T Tris-Buffered Saline and Tween 20

WEKA Waikato Environment for Knowledge Analysis

Chapter 1

Introduction

1.1 Motivation

Bringing a new therapeutic drug to market is expensive and time consuming. In the United States, bringing a single drug to market typically takes 10-15 years, and its cost can range from $800 million to $1 billion. For every one drug approved, there have been up to 10,000 drugs rejected and even after a drug has made it to market, there is still a potential for it to be removed [40].

Every year over 2 million people are affected by an adverse drug reaction (ADR) [24], out of which 26% can be attributed to drug-drug interactions (DDI) [29]. Elderly and cancer patient groups specifically, are at a higher risk from DDIs because they are administered multiple medications. Identifying DDIs early in the drug discovery and development process can reduce the number of hospitalizations due to drug interactions and allow more beneficial drugs to reach the market more quickly and at a lower cost.

1.2 Problem

Unfortunately, DDIs are difficult to test for due to their complex nature. Drugs are known to interact through a diverse array of mechanisms with the additional complication of affecting several different organs, tissues and pathways. This not only makes testing for all possible interactions difficult, but the prediction of DDIs is also often generalised to one mechanism of interaction, further limiting the discovery of new interactions.

Current computational methods in predicting DDIs are aimed at expediting and enhancing the drug discovery and development process and offer several approaches. Various computational approaches have shown to be successful, but similar to *in vitro* and *in vivo* methods are often limited to only one of the many possible ways that drugs can interact.

1.3 Hypothesis

It was hypothesized that computing could be used to build a classification model that predicts DDIs by analyzing and comparing the structure of drugs. Specifically, using a data set of molecular descriptors, computer modeling may be used to identify structural properties of drugs that lead to DDIs. Furthermore, it was hypothesized that pairs of drugs that were predicted to interact would lead to a change in cell death *in vitro*.

1.4 Objectives

A model based on the principle that structure determines function may present a solution that encompasses multiple drug interaction mechanisms by stripping underlying DDIs to the molecular level.

The objective of the present study was therefore to see if a quantitative structure-activity relationship (QSAR) classification model can be used to predict existing DDIs as well as new interactions between acetaminophen and select ingredients found in energy drinks by assessing cell death *in vitro*.

1.5 Contributions

This section introduces some of the shortcomings of current computational approaches to predicting DDIs and how a QSAR classification model can address them. Additionally an overview of *in vitro* model validation is presented.

1.5.1 QSAR modeling to predict drug-drug interactions

Many current computational approaches to predicting drug interactions are too specific to be applied to the broader spectrum of DDIs. Although considered successful, they are constrained to certain pathways and targets [44] or explicit mechanisms [51]. Some models rely on *in vitro* data as parameters [15], where the success of the model depends on the quality of the data.

By analysing fundamental properties of drugs at the molecular level, a QSAR classification model can predict DDIs that are not limited to any one organ, tissue or pathway.

The presented model also benefits from being validated *in vitro*, a crucial step that is often left out from the modeling process.

1.5.2 Acetaminophen and ingredients in energy drinks

Acetaminophen, also known as paracetamol, is commonly used for its analgesic (pain reliever) and antipyretic (fever reducer) effect and is metabolized by the liver. Acetaminophen is also associated with hepatotoxicity and is known to lead to cell death *in vitro* [7].

Several compounds are known to interact with acetaminophen with varying clinical significance ranging from major to minor. These compounds include agents found in energy drinks and other dietary supplements. Caffeine for example has been shown to increase the toxic potential of acetaminophen [7]. Studies have also suggested that compounds such as taurine, L-carnitine and creatine may have protective effects against cell toxicity [1, 41, 55]. Drug-drug interactions between acetaminophen and caffeine, creatine, pyridoxine, L-carnitine and taurine were evaluated, and in agreement with model predictions, no significant interactions were found.

1.6 Outline of thesis

The remainder of the chapters are organized as follows. QSAR modeling, an overview of classification modeling and DDIs are discussed in the next chapter. Predicting DDIs *in vitro*, *in vivo* and *in silico*, including related work is presented in Chapter 3. Chapter 4 describes the chemical descriptor based QSAR classification model, as well as *in vitro* validation. Results are outlined in Chapter 5, and Chapter 6 provides a summary and outlines future work.

Chapter 2

Background

2.1 Classification QSAR modeling

QSAR models are regression or classification models based on the assumption that structure determines activity. By using physico-chemical properties and theoretical molecular descriptors as predictor variables, they can summarize a relationship between the structure of a molecule and its biological activity [12].

From the structure of a molecule, numeric values describing the chemical composition, geometry, electronic distribution and connectivity of a molecule can be calculated [48]. Based on these numeric descriptors, molecules can be classified to a categorical value using a statistical classification algorithm or classifier [46].

QSAR models can not only be used to estimate the activity of new molecules but can also make predictions on the the activity of existing molecules [46].

2.2 Classification model using support vector machines

The problem of classification involves categorizing observations (samples or instances) by estimating a relationship between the features of the data and a categorical value

known as a class or a label. Building a classification model involves several steps and selection of appropriate algorithms.

2.2.1 Preprocessing

Preprocessing is an important step prior to classification. Raw data is often noisy and sometimes incomplete. Additionally there are several algorithms that are sensitive to scaling and require the features to be standardized so that they are normally distributed.

The relationship between observations and classes can also be clouded by features that provide no useful information or that do not help distinguish between classes. Furthermore, having too few features may hinder the performance of the classifier by overfitting the data, making it too specific to the training set.

Dimensionality reduction

Dimensionality reduction is performed to reduce the number of features in a data set. There are two forms: feature selection, used to select important features and feature extraction, used to map features to a lower dimensional space.

Feature selection Including associations with the classifier, there exist three feature selection techniques: filter, wrapper and embedded [26].

Filter method The filter method eliminates or includes features based on general properties of the data set [47]. For example, a feature may be eliminated if it is found to have zero variance across all samples. A feature such as this provides no useful information for distinguishing between classes. This method is not tailored to

any one classifier and for this reason may not choose the best features for a specific classifier. It is however a fast and simple approach [47].

Wrapper method Unlike the filter method, the wrapper method tends to find features suited to a specific classifier. When using the wrapper approach, a set of features is selected and then a classifier is evaluated [26]. This is repeated until the model is optimized and because of this can be computationally expensive [47].

Embedded method In the embedded method, feature selection is incorporated in the classifier and occurs during training. In a classifier such as decision trees, the feature that has the best ability to distinguish between classes is chosen in each stage of the algorithm [47].

Feature extraction There exist many feature extraction algorithms. Unlike feature selection, feature extraction preserves underlying feature information and transforms the data rather than trimming it.

Principal components analysis Principal components analysis (PCA) is one of many feature extraction techniques that is used to reduce the dimensionality of a data set. The original features are replaced by principal components, thats is, derived variables that are linear combinations of the original features. PCA benefits from its ability to retain most of the variability of the original features even when the number of components is much smaller than the original number of features [20].

2.2.2 Classification

Observations may be classified according to a set of instances with known labels (training set) or grouped according to patterns in the data set. Several algorithms exist for both approaches, each with varying performance depending on the nature of the data.

Unsupervised classification

The task of unsupervised classification involves finding patterns in data whose classification is unknown. Unsupervised classification includes algorithms that group or cluster data, separate independent variables from noise and generate probabilistic models.

K-means K-means is a fundamental unsupervised classification algorithm. It is often used as a basis for further assessment and much of its popularity can be attributed to its simplicity and speed [2].

The K-means algorithm begins by arbitrarily choosing k initial centers from a set of data points. The remaining data points are then assigned to the center that is closest to it, forming clusters. The centers of each cluster are recalculated after each new addition, as the mean of all the points in that cluster.

Supervised classification

Supervised classification involves estimating a relationship between features and labels from training data whose classification is known. The estimated relationship can then be used to classify new, unseen observations.

Support vector machines Support vector machines (SVM) is a supervised classification method that has the advantage of being memory efficient and still effective in analysing high dimensional data [38]. SVM also performs well when given sparse or noisy data [10].

SVM performs classification by generating hyper-planes to separate binary labeled training data. Here, SVM benefits from being versatile, allowing for the use of different kernel functions. Kernel functions are useful in classification when the relationship between the data and classes is not necessarily linear; kernels allow for a non-linear mapping to feature space [10].

2.3 Multi-label classification

Multi-label classification is a classification problem where each observation may belong to more than one label or class that are often correlated [47]. A multi-class problem differs from a multi-label problem in that its classes are disjoint; a fruit may be classified as an apple or a banana but it cannot be both. An example of a multi-label data set is shown in Figure 2.1 on the next page. Multi-label classification techniques can be divided into two main categories: problem transformation and algorithm adaptation [50].

Day	Sunny	Warm	Raining
1	1	1	0
2	0	0	1
3	0	1	0

Figure 2.1: A multi-label data set describing weather. The membership of an observation (Day) to a label is indicated with a 1. The labels for each observation corresponds to the indices of its membership (starting from 0). For example, Observation 1 would have labels 0 and 1, Observation 2 would have label 2 and so on.

2.3.1 Problem transformation

Problem transformation methods transform the multi-label problem either into several binary classification problems or a single multi-class problem [50]. There are a multitude of algorithms that exist for both approaches, two of which are described below.

Binary relevance

The binary relevance, also known as the one-vs.-rest (OvR) approach, transforms a multi-label problem into several binary classification problems. This is done by separating the data set according to labels and building independent classifiers for each of these subsets. However, because only one label is considered at a time, the correlation among labels is not accounted for [47].

Label powerset

Unlike the binary relevance approach, the label powerset method does consider the correlation among labels. A multi-class data set is created by making a single class from each unique combination of labels. The main shortcoming of this method occurs when some class values are only represented by a small number of observations, creating a multi-class data set that is unbalanced [47].

2.3.2 Algorithm adaptation

Algorithm adaptation methods are classification algorithms that are extended to handle multi-label data. There are several classification algorithms that have been adapted by incorporating problem transformation methods, such as binary relevance, into the classification algorithm [47].

2.4 Drug-drug interactions

A DDI may be defined as any drug activity that is either greater or less than expected in the presence of one or more other drugs [13]. DDIs can be classified in two categories: pharmacodynamic interactions - what the drug does to the body, and pharmacokinetic interactions - what the body does to the drug. Drugs can interact at the pharmacodynamic level at the receptor, during signal transduction or through separate physiological processes. Absorption, distribution, metabolism, excretion and transport (ADMET) are factors affected by pharmacokinetic interactions [17, 4].

2.4.1 Pharmacodynamic

How the body responds to a drug is complex and can occur through a variety of mechanisms. Pharmacodynamic interactions can be summarized as follows.

Pharmacological receptors

At the receptor, a drug can either be an agonist or an antagonist. An agonist will increase the effect of another drug whereas an antagonist will block the effect of another drug [56].

Agonists A pure agonist, for example drug A, will bind to the same receptor as drug B, leading to an overall increase in drug effect (refer to Figure 2.2 on the next page). The increase in effect may either be additive (the sum of the individual effects), synergistic (greater than the sum of the individual effects) or potentiated (drug A will increase the effect of drug B, but drug A is devoid of the effect of drug B) [37]. There are also partial agonists, where drug A will bind to the same receptor as drug B, but cause less of an effect than a pure agonist [56].

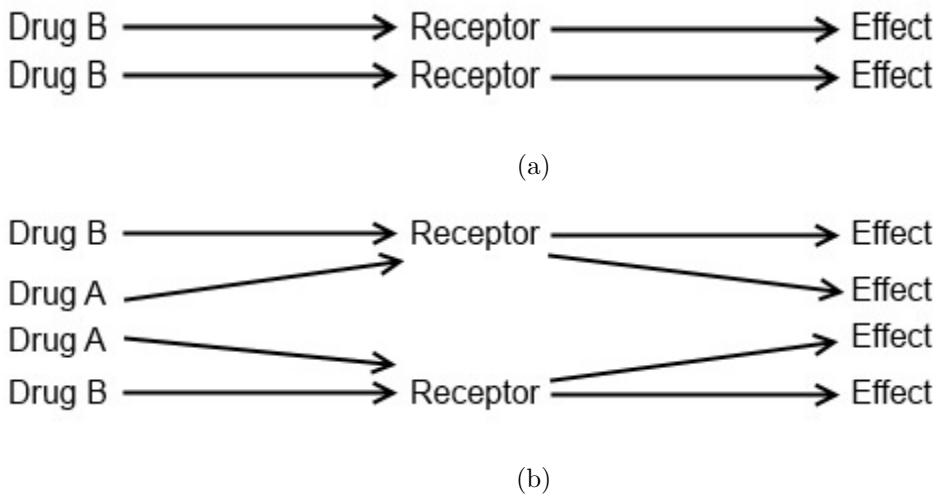


Figure 2.2: The effect of a receptor agonist. Figure 2.2(a): The expected effect of drug B. Figure 2.2(b): The increased effect of drug B caused by drug A.

Antagonists Antagonists may either be competitive or non-competitive. Given that drug A is a competitive antagonist, it will compete for the same receptors as drug B, leading to a decrease in effect of drug B. However if drug A is a non-competitive antagonist, it will bind to the receptor regardless the presence of drug B, and drug B will no longer have an effect at the receptor [4] (see Figure 2.3 on the next page).

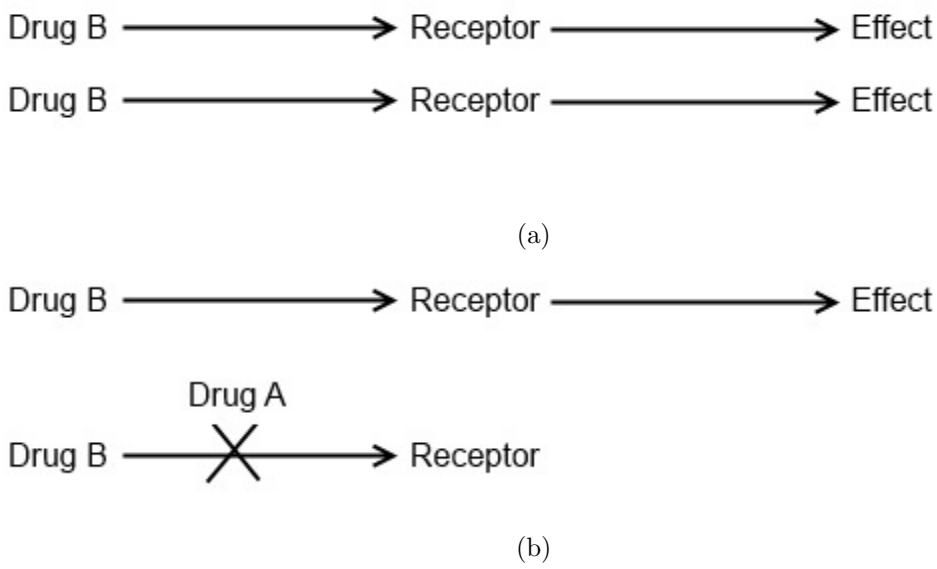
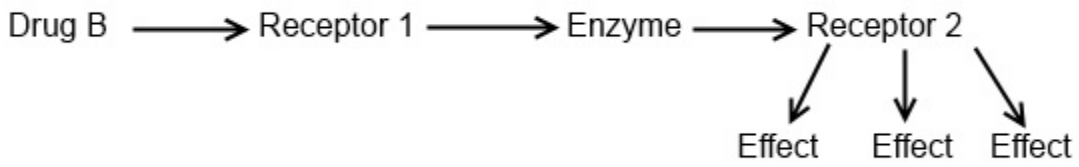


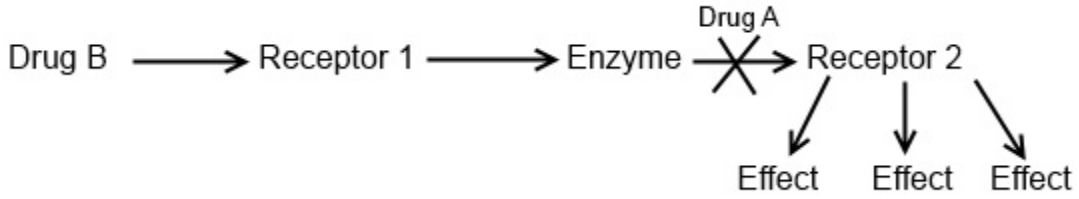
Figure 2.3: The effect of a receptor antagonist. Figure 2.3(a): The expected effect of drug B. Figure 2.3(b): The decreased effect of drug B caused by drug A.

Signal transduction

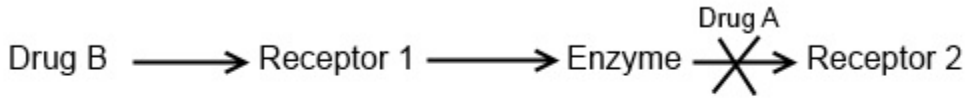
During signal transduction, drug A may affect the signal caused by drug B [4]. For example, drug B binds to receptor 1, causing the release of an enzyme that binds to receptor 2 that causes an effect. If drug A is introduced to the system and it binds to receptor 2 thereby inhibiting the binding of the enzyme, then the effect caused by the binding of the enzyme to receptor 2 would also be inhibited (see Figure 2.4).



(a)



(b)



(c)

Figure 2.4: The effect of drug A on the signal transduction of drug B. Figure 2.4(a): Normal signal transduction. Figure 2.4(b): Drug A binds to receptor 2. Figure 2.4(c): The binding of drug A inhibits the binding of the enzyme and the effect is inhibited.

Physiological processes

Physiological DDIs are interactions that occur through separate physiological processes; a physiological antagonist will have the opposite effect of a drug but through separate mechanisms [37]. For example, histamine is known to decrease blood pressure by binding to histamine receptors in blood vessels. On the other hand, epinephrine

increases blood pressure by binding to adrenoreceptors in blood vessels [53]. The two events occur through separate pathways but can still in this way affect one another.

2.4.2 Pharmacokinetic

The effect of a drug can be modified by changing its absorption, distribution, metabolism, excretion and transport. DDIs that occur by affecting these ADMET factors change the normal behaviour of a drug by altering how long it stays in the body and its expected concentration.

Absorption

Decreasing the time a drug is in the body decreases the time the body has to absorb the drug. This can occur for example, by changing gastrointestinal motility through the use of laxatives or diuretics that can cause the rapid excretion of other drugs that are in the system. Drugs may also interact by altering gastric pH [17]; ranitidine, a drug used to treat and prevent symptoms associated with too much acid production in the stomach, increases gastrointestinal pH by decreasing stomach acid production. This may result in the increased absorption of basic drugs such as triazolam [35]. Some drugs when taken together can cause the formation of non-absorbable complexes through binding or chelation. The body is unable to absorb these complexes and they are excreted [9].

Distribution and transport

Many drugs that are taken are highly bound to plasma proteins in the blood to be transported in the body. Drugs in this form are not absorbed [37]. However, the

presence of a second drug may cause the plasma proteins to release the first drug. This occurs when the second drug has a higher affinity for the plasma proteins. It can lead to an increased concentration of the first drug above the intended dose and cause toxicity [17].

Metabolism

Metabolism plays a prominent role in how drugs behave in the body. Cytochrome P450 enzymes (P450s) are largely responsible for drug metabolism. They are accountable for the metabolism of approximately 75% of the drugs that enter the human body which usually occurs in the liver [14]. There are over 50 known P450s that metabolize different chemicals in different ways [28]. The 5 major P450s are: CYP1A2, CYP2C9, CYP2C19, CYP2D6 and CYP3A4, with 80-90% of P450 metabolism accounted for by CYP3A4 and CYP2D6 [31].

20-30% of ADRs that occur because of DDIs are P450 related and are caused by either the induction or inhibition of P450s [31]. Drugs may either increase (inducers) or decrease (inhibitors) P450 activity. Induction occurs when certain substrates increase the rate of P450 synthesis or reduce the rate of P450 degradation [5]. The induction of P450s may lead to an increase in metabolic rate and consequently a decrease in drug concentration and effect. This is known to be caused by compounds such as polycyclic aromatic hydrocarbons, chemicals that are found in cigarette smoke, which can induce CYP1A2 [3, 6]. Induction of P450s may however also lead to an increase in therapeutic effect, by increasing the metabolism of a pro-drug to its active form. The analgesic effects of codeine for example may be increased by inducing its metabolism to morphine by CYP2D6 [3].

In contrast, inhibition of P450s leads to a decrease in metabolism and a consequent increase in drug concentration to possibly toxic levels [31]. This may result from the binding of certain drugs to P450s or their intermediates, rendering them inactive [5], and is known to occur by certain chemicals that can be found in grapefruit juice [3]. Furanocoumarins, organic compounds found in grapefruit juice, can increase drug bioavailability and enhance the beneficial or adverse effects of several medications by inhibiting CYP3A4 in the gut, thereby increasing drug absorption [3, 21].

P450 related toxicity may also be attributed to toxic metabolites that arise from the binding of substrates [31]. The hepatotoxicity of acetaminophen for example can be attributed to N-acetyl-p-benzoquinone-imine (NAPQI), which is produced from the metabolism of acetaminophen by P450s [7] (Figure 2.5). Studies of human liver cells suggest that CYP2E1 and CYP2A6 are the most responsible for the formation of NAPQI [54]. Hepatotoxicity is known to result from the interaction of acetaminophen with caffeine, which is suspected of enhancing the metabolism of acetaminophen to NAPQI [7].

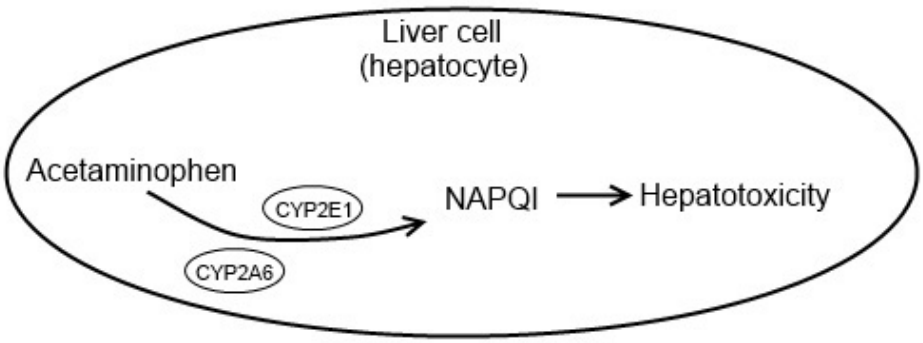


Figure 2.5: Acetaminophen is metabolised by P450s such as CYP2E1 and CYP2A6, forming NAPQI and resulting in hepatotoxicity.

Excretion

A drug may increase or decrease the excretion of another drug by altering urine pH. Increasing the pH of urine (making it more basic), increases the excretion of acidic drugs. On the other hand, decreasing the pH of urine (making it more acidic), increases the excretion of alkaline drugs. This may occur through mechanisms of diffusion, where the drug moves down the concentration gradient or by affecting transport proteins [17].

Chapter 3

Predicting Drug-Drug Interactions

Evaluating the potential DDIs of a new chemical entity (NCE) is a crucial step in drug discovery and design. Several methods have been developed *in vitro* as an initial step to further guide assessment *in vivo*. Additionally, *in silico* methods exist that may supplement the screening and early detection of DDIs.

3.1 *In vitro* evaluation to predict metabolism-mediated drug-drug interactions *in vivo*

Predicting P450-mediated interactions often begins *in vitro* by identifying P450 substrates, inducers and inhibitors using human liver tissues, cells or microsomes that express specific enzymes. Generally, the chance of a DDI occurring increases if the compound has a high affinity for a single P450 as opposed to many [57]. If *in vitro* results suggest that a DDI is likely to occur, further studies are conducted *in vivo*, using model organisms such as mice, before advancing to clinical trials [40].

3.1.1 Predicting P450 substrates

The selection of interacting drugs *in vitro* is important for proper evaluation *in vivo*. A drug interaction is likely to occur between a NCE and an inducer or inhibitor if it contributes to greater than 25% of the total clearance (the rate of elimination) of the NCE [57]. Experiments are performed to identify the enzymes involved in metabolising a NCE and to explore its metabolic pathway. If it is found that the NCE is likely to be an inducer or inhibitor of a P450 enzyme, then liver microsomes may be used to determine specific P450s involved. Further studies may be conducted *in vivo* if warranted, to see if the expected drug response is altered by inducers or inhibitors [57].

3.1.2 Predicting P450 inducers

When a NCE is an inducer of a P450 enzyme, it increases the clearance of that enzyme's substrates. By using known substrates of P450 enzymes, enzyme activity can be evaluated by measuring metabolite concentration. If the NCE causes an increase in enzyme activity above expected levels, then it is considered to be an inducer of that enzyme [57].

3.1.3 Predicting P450 inhibitors

Similar to predicting P450 inducers, known P450 substrates are also used to predict inhibitors. Using liver cells expressing a P450 enzyme, a NCE is administered with known substrates of the P450. If the NCE is an inhibitor of the P450, it will inhibit its metabolism, leading to an increased substrate concentration that is above the expected amount [57].

3.2 *In silico* modeling to predict drug-drug interactions

There have been several computational approaches to predicting DDIs [8, 13, 16, 18, 19, 25, 27, 31, 32, 39, 44, 49, 51, 52]. Although unique, many of the models include methods that overlap and are based on similar underlying principles. P450 metabolism, molecular structure and mechanisms are common themes among models.

3.2.1 Classification

Classification based methods are commonly seen in predictive modeling. For example, classification algorithms have been used to predict DDIs by predicting P450 substrate specificities. Various techniques exist that range in accuracy from 67-89% [31]. Examples of such techniques include classifying various drugs and their analogs using SVM, decision trees (DT) and using WEKA (Waikato Environment for Knowledge Analysis) for feature selection [49]. Among other approaches are the use of several algorithms such as SVM, random forest (RF), kappa nearest neighbor (kNN) and DT to classify P450 inhibitors and non-inhibitors [51]. WEKA has also been used to build models using various classification algorithms to identify substrate specificities for P450s [32].

3.2.2 Data mining

Data mining algorithms are used to discover statistically strong associations in data [16]. Text mining (data mining of text) has been shown to be particularly advantageous in deriving information useful in predicting DDIs.

Statistical mining

In 2010, a method using text mining was developed to find drug interactions by identifying associations between drugs and adverse events in the Food and Drug Administration (FDA) Adverse Event Reporting System (AERS) [16]. A spontaneous reporting system (SRS) such as the AERS is a database of reports of suspected ADRs. By using a data mining algorithm to look at associations including at least two drugs and one ADR, the method identified 1,868 potential drug interaction adverse effects. Out of the drug interactions identified, 35% of drug pairs are known to interact, the remaining 65% are unknown interactions.

Literature discovery

A 2012 study showed that information taken from literature can be correlated to medical record databases and used to predict novel DDIs [8]. A text mining approach was developed to identify drugs from published *in vitro* pharmacology experiments that were either P450 substrates or inhibitors. Based on this mined information, 13,197 potential DDIs were predicted. Using a clinical repository, this predicted set was reduced to 3,670 drug pairs that are actually taken by patients. Medical records were mined further to identify drug pairs that appeared to be correlated with symptoms of a DDI.

3.2.3 Descriptor-based

Many prediction techniques have made use of various descriptors to predict DDIs. They have shown not only that structure is an important determinant of function, but also that descriptors are versatile in modeling the diverse mechanisms of DDIs.

Predicting binding affinities

In 2008, a study was done on predicting DDIs based on affinity models developed for CYP2C9 [19]. The study was based on the general rule for DDIs that if a compound's binding affinity for a P450 enzyme is greater than 10μM, then it is unlikely to cause important clinical DDIs. Hudelson and colleagues recognised however that different modeling techniques, models based on crystal structures or QSAR models for example, each have their own shortcomings. As a result they used four different methods, two based on descriptions of global structure, a third relating these descriptions to activity and the fourth using a graph-theoretic approach to predict activity from molecular structure. When these four methods agreed, the predictive accuracy was 94% (based on a training set of 276 compounds and a diverse validation set of 50 compounds).

Simulating liver cells

A novel approach to predicting DDIs was introduced by Sheikh-Bahaei and colleagues in 2011. They took on the ambitious task of creating computer programs to simulate hepatocyte cultures. Their simplified virtual liver cells included the main drug metabolizing components of cells such as transporters and P450 enzymes CYP3A4, CYP2D6, CYP2C9, CYP1A2, CYP2C19 and CYP2E1. They also created objects that simulated 73 compounds using molecular descriptors and physico-chemical properties, then used probabilities to estimate their metabolism by enzymes. A set of rules was established to predict how chemical and physical properties of drugs would influence the probability of certain events. The model was able to predict clearance rates accurately 79% of the time [44].

QSAR modeling

QSAR models have been widely used to predict drug metabolism through various statistical techniques and molecular descriptors [25]. Because QSAR models are based on molecular descriptors, they are useful in predicting interactions that arise because of the chemical structure of molecules, as in metabolism. As outlined by Li *et al.*, QSAR models have been used to predict P450 metabolism, clearance and metabolic stability. Although ranging in success, recommendations have been outlined that address the lack of sufficient interpretability and poor predictability for novel drugs of many existing QSAR models. These recommendations include the importance of proper division of the training and test sets, choosing appropriate end points, and the generation of rationally designed molecular descriptors [25].

3.2.4 Similarity measure

Defining a similarity measure between drugs has been shown to be beneficial in modeling DDIs. Comparing drugs by looking at shared features can predict novel DDIs, and additionally reveal information about underlying mechanisms and effects.

Inferring drug interactions

Gottlieb *et al.* developed a method that performed well in predicting pharmacokinetic, P450 related DDIs and pharmacodynamic, non-P450 associated DDIs [13]. They assembled a database that distinguished between P450-related DDIs, potential P450-related DDIs and non-P450-related DDIs. Given this known set of DDIs, they constructed features based on similarity measures and then applied a classifier to predict new DDIs. Seven similarity measures were used including chemical similarity,

ligand-based chemical similarity, similarities based on side-effects and similarities constructed between drug targets. Their method was not only able to predict potential DDIs but also DDI recommendations and related P450s.

Interaction profile fingerprints

In a 2013 study, a novel approach was presented that predicted DDIs based on the similarity of profiles representing drug interactions for a particular drug [52]. These profiles, coined interaction profile fingerprints (IPF), associate a drug with a number so that the interactions for a drug could be coded as a vector. IPFs were generated from 928 drugs and a set of 9,454 unique DDIs. Predictions of new DDIs were made from the calculations of a similarity score defined as the ratio between the number of interactions shared by two IPFs. 40-50% of DDI predictions made were correct, and by referring back to true DDIs, hypotheses on the resulting ADRs could be made.

Predicting target-ligand interactions

Recently, Lounkine and colleagues addressed the daunting task of predicting off-target ADRs by creating an association metric and using a similarity ensemble approach (SEA) [27]. SEA calculates the likelihood a ligand will bind to a target based on the chemical similarity it shares with targets that have known ligands. Using a statistical model to account for random similarity and by creating a drug-target-adverse drug reaction network, they were able to predict half of the known ADRs that were tested.

3.2.5 Network-based

Networks are useful in highlighting patterns in data. By representing biological components as nodes and their relations as edges, networks can be used to elucidate potential DDIs and mechanisms.

Drug-gene interactions

Percha *et al.* proposed a text mining method that extracts interactions from literature and attempts to predict DDIs using a classifier [39]. With an accuracy of approximately 80%, Percha and colleagues were able to hypothesize on not only whether an interaction will occur, but also on its mechanism. Drug-drug interactions are often the result of two or more drugs interacting with similar pathways or genes. These drug-gene interactions are dispersed across literature and generally unknown. To this problem Percha and colleagues offered an interesting solution. By searching various abstracts for drug names, genes and specific context words, they were able to construct a complex network that described the relationships between genes and drugs. All the connecting relations and context terms along the shortest path between each pair of drugs (the path from one drug to a gene to a second drug) could then be used as descriptor values for a classifier. Percha and colleagues proceeded by training a random forest classifier. For each interacting pair, the paths that were voted as "interacting" the most by the classifier were predicted to be its means of interaction.

Protein-protein interactions

The use of a protein-protein interaction network considers DDIs in the context of molecular systems [18]. Using known drug targets, Huang *et al.* showed that pharmacodynamic DDIs could be predicted by looking at connections between possible drug pairs. They found that the average distance of known DDI targets is significantly shorter than the average distance of all possible drug pairs in the network. Using this information, a metric that measured the strength of a connection between drug targets and potential DDIs was developed. The method was able to predict 9,626 potential DDIs with an accuracy of 82%. Through the use of a protein-protein interaction network, the algorithm also provided insight on potential molecular mechanisms and physiological effects of the predicted DDIs.

3.2.6 QSAR modeling to predict drug-drug interactions between acetaminophen and ingredients in energy drinks

In this thesis work, a QSAR model was developed to predict DDIs and address some of the shortcomings of existing methods that have been developed. Limitations such as the predictability and interpretability of existing QSAR models, the ability of models to predict DDIs mediated through only one mechanism or one P450 and relying on *in vitro* information prior to modeling are some of the issues that the developed QSAR model aimed to address. The model, unlike many existing approaches, also benefits from being validated *in vitro* as outlined in the following Chapter.

Chapter 4

Materials and Methods

An overview of the model design process can be seen in Figure 4.1 on the next page and is described in detail as follows.

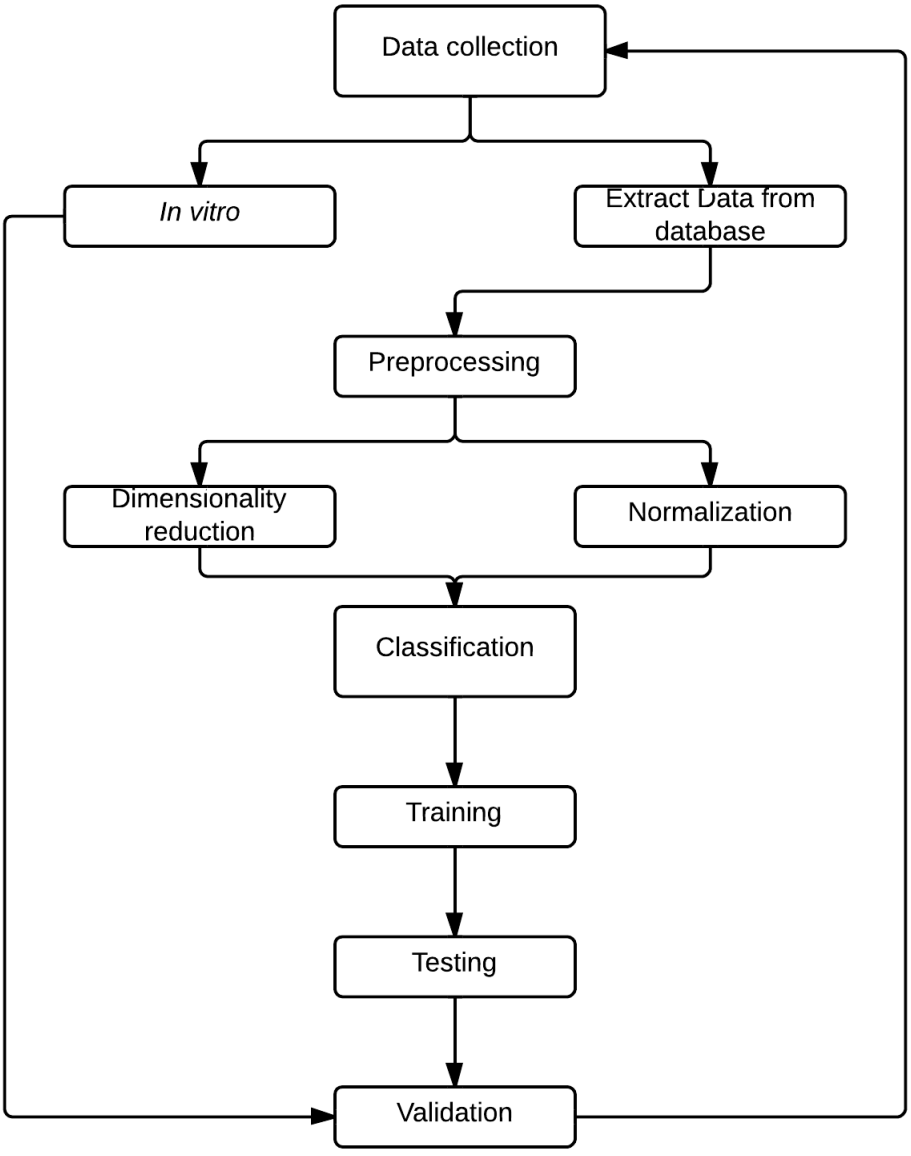


Figure 4.1: Flowchart outlining the model design process. Data collection can be divided into two steps: *In vitro* analysis and data extraction for model construction. The combined final results of the model can be used to further model improvement.

4.1 QSAR modeling

SciKit Learn [38] was used to implement all computational modeling algorithms (PCA, OvR-SVM and K-means).

4.1.1 Data collection

In addition to data collection described in the following paragraphs, *in vitro* data collection is also outlined in Section 4.2.

Data set construction

A data set of 6,853 entries of both FDA approved and experimental drugs combining chemical, pharmacological and pharmaceutical data, as well as drug target information was obtained from DrugBank [23]. Pybel [34] was used to extract a list of generic names and their corresponding drug interactions. Drugs listed as interacting that were not part of the initial 6,853 data set were excluded, resulting in a set of 6,583 drugs and their interactions. Interactions were stored as a matrix, where the rows and columns represent drugs, and a 1 denotes an interaction, 0 otherwise. It is important to note that a 0 in the interactions matrix does not necessarily indicate that there is no interaction, only that there is no known interaction (see Figure 4.2 on the next page). This generated a set of 42,646,653 non-interacting drugs pairs and 20,371 interacting drug pairs.

	Drug A	Drug B	Drug C
Drug A	0	1	1
Drug B	1	0	1
Drug C	1	1	0

Figure 4.2: A 1 denotes an interaction between drugs, 0 indicates that there is no known interaction. When treated as a multi-label classification problem, the labels for each drug would be the indices (starting from 0) of its interacting drug. For example, Drug A would have labels 1 and 2, Drug B would have labels 0 and 2 and Drug C would have labels 0 and 1.

QSAR descriptor calculation

Cinfony [33] was used to access the Chemistry Development Kit (CDK) [48] to calculate 292 QSAR descriptors. These physico-chemical and theoretical molecular descriptors could then be used as features for all 6,532 drugs. A full list of QSAR descriptors can be found in Appendix A.1

4.1.2 Preprocessing

The data was preprocessed prior to modeling through the imputation of missing values using the median of the column (feature) of the missing value. The data was also standardized (Gaussian with zero mean and unit variance).

Dimensionality reduction

Prior to feature extraction using PCA, features were filtered by removing any column that did not vary at all or that varied above the max allowance (set to 99), where:

$$max\ allowance = \frac{\#distinct\ values}{\#total\ values} \times 100 \tag{4.1}$$

This filter eliminated 92 features, resulting in a data set of 6,532 rows and 200 columns. PCA was performed using the method proposed by Thomas P. Minka: "Automatic choice of dimensionality using PCA" [30], which reduced the data set further by only one attribute.

4.1.3 Classification

SVM

Although SciKit Learn [38] has a limited number of algorithms implemented that allow for multi-label classification, SVM was found to be well suited to the data set. SVM was chosen for classification because of its ability to handle sparse and noisy data, perform well in high dimensional space and because of its memory efficiency. Multi-label SVM, implementing the OvR method, was used for supervised classification and classification was performed using a polynomial, radial basis function (RBF) and linear kernel. The use of OvR-SVM with a linear kernel yielded the highest number of correct predictions compared to the other kernels, and was therefore used in subsequent modeling. Further modeling was done by building classifiers for smaller subsets of the data and having these smaller models vote on classification. Unique sets of 195, 390, 780, 1,500 and 3,000 observations were randomly chosen

from the complete data set of 6,532 drugs, and a classifier was built for each set until all observations were chosen at least once. Additional classification was executed using weighted, linear OvR-SVM that assigned weights to the class labels for equal representation.

K-means

K-means clustering was performed using $k = 2$, 3, 5, 7 and 10 clusters. Additional unsupervised methods were also used yielding similar results, however K-means was able to separate the data the best and the most consistently.

4.1.4 Model evaluation

SVM

Results for each classifier were averaged over 100 runs and stored. Classifier performances were then evaluated using a confusion matrix displaying correct and incorrect predictions. Predicted interactions for acetaminophen, pyridoxine, creatine, L-carnitine, taurine and caffeine were ranked from highest to lowest based on the number of times they were predicted out of the 100 runs. DrugBank [23] was then used to analyse the interactions by comparing common targets, enzymes, and indications.

K-means

Resulting cluster assignments from K-means classification were first evaluated through cluster analysis and then by constructing a DDI network in Cytoscape [42]. Cluster analysis was done for each k by calculating a score as follows: for each drug in a cluster, find the distribution of its interacting drugs among all the clusters. By evaluating the

distribution of the members among the clusters, it was determined that using k=2 clustered the data the best and was subsequently used in further evaluation.

Using known DDIs, drugs could be represented as nodes and interactions as edges in a network. The network was visualised using the Spring Embedded layout, where nodes that are more connected are held closer together as if by springs. Cluster assignments were then overlaid on the network as color. Further evaluation was done by entering P450 drug metabolism manually using the "P450 Drug Interaction Table" from the Indiana University Division of Clinical Pharmacology [36]. A total of 386 P450 labels were used, and drugs were labeled as CYP1A2, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4, CYP3A5, CYP3A7 substrates, inducers or inhibitors. This information was then imported as node attributes into Cytoscape and clusters were analysed; the following information was found:

% **Cluster** the proportion of all the P450s in the cluster occupied by each individual P450

% **P450** the proportion of the individual P450 that is in the cluster

4.2 *In vitro* model validation

4.2.1 Materials

Hepa 1-6 (mouse liver hepatoma) cells were purchased from the American Type Culture Collections (ATCC, Manassas, Virginia). Fetal bovine serum (FBS) was obtained from Fisher Scientific Ltd. (Ottawa, ON) and remaining cell culture media reagents were acquired from Life Technologies (Burlington, ON).

4.2.2 Cell culture

Hepa 1-6 cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 5% penicillin-streptomycin and 10% FBS. Cells were maintained at 5% CO_2 and 37°C.

4.2.3 Transfection

Hepa 1-6 cells were seeded in 1ml complete media in 6-well plates at a concentration of 5.0×10^5 cells per well and allowed to incubate until 50-80% confluence was reached. CYP2E1 expression plasmid was obtained from Dr. Arthur Cederbaum, Mount Sinai School of Medicine (New York, NY). 2.5 µg of DNA and either 5, 8 or 10 µl of Lipofectamine™(Life Technologies) was added to each well, in addition to enough Opti-MEM™(Life Technologies) to bring the total volume to 1ml. Additional transfections were performed by seeding in different sized plates at various densities and amounts of Lipofectamine™. Several other variables, such as DNA:transfection reagent ratio, incubation times and cell confluence, were also manipulated.

Western blotting

Cells were washed with phosphate buffered saline (PBS) and collected. The Bio-Rad Laboratories Bradford Protein Assay (Mississauga, ON) and an Ultrospec 3100 Pro scanning spectrophotometer (Biochrom Ltd., UK) were used to determine protein concentration. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate 30 µl of each sample which were then transferred. Once transferred, the membrane was rinsed with a mixture of Tris-buffered saline and Tween

20 (TBS-T) and then incubated in a block of 3% milk and 3% bovine serum albumin (BSA) in TBS-T for 30 minutes. Rabbit polyclonal anti-Cytochrome P450 2E1 antibody (ABCAM, Toronto, ON) was diluted in block at a ratio of 1:1,000, and incubated with the membrane overnight at 4°C. The membrane was then washed with TBS-T and incubated with anti-rabbit secondary antibody (ABCAM), diluted in block at a ratio of 1:10,000, at room temperature for 1 hour. The membrane was once again washed with TBS-T and then developed using Western Lightning *Plus* ECL (Perkin Elmer, Waltham, MA).

4.2.4 Exposure

Hepa 1-6 cells were seeded in 24-well plates at a density of 1.0×10^5 cells per well in 500 µl of media and incubated overnight. Acetaminophen, taurine, caffeine, L-carnitine, creatine, pyridoxine and theophylline (Sigma-Aldrich Chemical Co.) were dissolved in 1% reduced serum media at the following concentrations:

Acetaminophen 250 µg/ml as per [43]

Taurine 1800 µg/ml as per [55]

Caffeine 485 µg/ml as per [7]

L-carnitine 16 µg/ml as per [41]

Creatine 657 µg/ml as per [1]

Pyridoxine 846 µg/ml as per [11]

Theophylline 300 µg/ml as per [45]

500 µl of each solution was added to separate wells in duplicate. 500 µl of a solution of acetaminophen and of either taurine, caffeine, L-carnitine, creatine, pyridoxine or theophylline solutions were also added to separate wells in duplicate. Theophylline, a top predicted interaction of both pyridoxine and caffeine, was used to validate predictions made by the QSAR model. Additional controls used include cells in media with no treatment and media alone.

4.2.5 Cytotoxicity

Cells were collected after 24 hours exposure and cell death was assessed using a lactate dehydrogenase (LDH) assay and an Ultrospec 3100 Pro scanning spectrophotometer (Biochrom Ltd.). A 0.1 M phosphate buffer (PB) with a pH of 7.4 was used in the following preparations:

Pyruvate solution Pyruvic acid (Sigma-Aldrich Chemical Co.) at 0.0127g/10ml PB

β-NADH solution β-Nicotinamide adenine dinucleotide (Sigma-Aldrich Chemical Co.) at 0.010g/ml PB

800 µl of PB was pre-mixed with 200 µl of pyruvate solution prior to reading. 200 µl of sample and 100 µl of β-NADH were mixed with the PB/pyruvate solution by pipetting and inversion and read immediately. Absorbance was read at 340 nm for 1 minute. Absorbance at the start and at the end of the readings, as well as the slope, were recorded.

4.2.6 Data analysis

Average percent cell viability was calculated as follows:

$$\% \, viability = 1 - \frac{slope_{media}}{slope_{media} + slope_{lysate}} \times 100 \tag{4.2}$$

Analysis was done using a two-way analysis of variance (ANOVA) with a significance of $p<0.05$.

Chapter 5

Results

5.1 Model evaluation

Confusion matrices were calculated to evaluate QSAR model performances. Based on the results, the best classifier was chosen and results were analysed further using DrugBank [23]. Additional validation was performed *in vitro* and analysed by evaluating differences in average percent viability.

5.1.1 SVM

It was determined that a weighted OvR-SVM classifier with a linear kernel had the best performance; model accuracy was found to be 99%, predicting approximately half of known interactions, as can be seen in Figure 5.1 on the next page where:

Accuracy is the proportion of the total number of predictions that were correct

Positive predictive value is the proportion of interactions that were correctly predicted by the model out of all interactions predicted

Negative predictive value is the proportion of non-interactions that were correctly predicted by the model out of all non-interactions predicted

Sensitivity is the proportion of interactions that were correctly predicted by the model out of all actual interactions

Specificity is the proportion of non-interactions that were correctly predicted by the model out of all actual non-interactions

Model specificity increased from 25% when using non-weighted OvR-SVM classifier to 49% when adding a weight assignment to labels during classification. Additional confusion matrix results are summarized in Table 5.2 for comparison.

		Target		
		Positive	Negative	
Model	Positive	3314	60377	Positive predictive value = 5%
	Negative	3465	4676528	Negative predictive value = 100%
		Sensitivity = 49%	Specificity = 100%	Accuracy = 99%

Figure 5.1: The calculated confusion matrix, averaged over 100 runs, for a weighted SVM classifier with a linear kernel. From the sensitivity it can be seen that the model predicts approximately half of known interactions and almost all non-interacting drug pairs are predicted as seen from specificity.

Kernel	Linear							RBF	Polynomial
Number of samples	6532	6532, WEIGHTED	1500	780	390	195	3000	6532	6532
Accuracy	100%	99%	100%	100%	100%	100%	100%	100%	100%
Number of true positives	728	3314	939	843	713	592	976	65	244
Number of true negatives	4733011	4676528	37837045	38068971	38412484	38547270	36316750	4736780	4735883
Number of false positives	3785	60377	70794	76414	71870	63570	45924	40	1052
Number false negatives	6160	3465	18224	18551	19013	19268	17813	6800	6505
Positive predictive value	16%	5%	1%	1%	1%	1%	2%	66%	19%
Negative predictive value	100%	100%	100%	100%	100%	100%	100%	100%	100%
Sensitivity	11%	49%	5%	4%	4%	3%	5%	1%	4%
Specificity	100%	99%	100%	100%	100%	100%	100%	100%	100%

Figure 5.2: Summary of calculated confusion matrices.

The model did not predict any interactions between acetaminophen and select ingredients found in energy drinks. All predicted interactions for pyridoxine, creatine, L-carnitine and taurine are not previously known. The model appears to be a better predictor of interactions for caffeine than for acetaminophen; 8 of 13 interactions were correctly predicted for caffeine 47-100% of the time, and only 2 of 6 interactions were correctly predicted for acetaminophen 13-56% of the time (refer to Table 5.1 on the next page). Lithium, which is known to interact with caffeine, was predicted to interact with caffeine 100% of the time. Top predicted interactions are defined as the the predicted interactions that are ranked higher than the average prediction rankings for the specific drug. Top predicted interactions are shown in 5.1 on the next page. True interactions are shown in bold. A list of all known interactions for each of acetaminophen, pyridoxine, creatine, L-carnitine, taurine and caffeine can be found in Table 5.2.

Acetaminophen	%	Pyridoxine	%	Creatine	%	L-carnitine	%	Taurine	%	Caffeine	%	Pyridoxine	%	Caffeine	%
Timolol	56	Verapamil	100	Labetalol	86	Betaxolol	74	Montelukast	68	**Lithium**	100	Acetohexamide	63	Primidone	61
Trandolapril	41	Nadolol	100	Pargyline	77	Clarithromycin	21	Thioguanine	56	Disulfiram	97	Isocarboxazid	59	Pefloxacin	61
Anisindione	38	Aminophylline	100	Didanosine	74			Labetalol	38	Isoniazid	95	Desipramine	59	**Regadenoson**	61
Treprostinil	34	Esmolol	100	Lamivudine	71			Azilsartan medoxomil	29	**Adenosine**	92	Torasemide	56	Rofecoxib	61
Dicumarol	31	Theophylline	100	Glucosamine	71			Torasemide	24	Acyclovir	92	Mephenytoin	56	Halothane	61
Phenytoin	25	Timolol	100	Timolol	66			Trandolapril	21	Pentoxifylline	92	Glycodiazine	56	Amprenavir	58
Methotrexate	25	Warfarin	100	Oxprenolol	63			Nateglinide	18	Temazepam	92	Trimipramine	50	Phenobarbital	55
Sertraline	13	Oxprenolol	100	Penbutolol	60			Chlorothiazide	18	Carbamazepine	92	Primidone	47	Quinidine	55
Lurasidone	13	Bisoprolol	100	Nadolol	57					Pindolol	89	Betamethasone	47	Mephenytoin	55
		Dyphylline	97	Treprostinil	57					Propranolol	89	Piroxicam	47	Bromazepam	53
		Oxtriphylline	97	Methylphenidate	57					Nadolol	87	Tamsulosin	44	Sitaxentan	53
		Acenocoumarol	97	Betaxolol	54					Aminophylline	87	Cortisone acetate	44	Darunavir	50
		Propranolol	97	Atenolol	54					Acenocoumarol	87	Salmeterol	44	Rifampin	50
		Tolbutamide	97	Pindolol	54					Timolol	87	Methylprednisolone	44	Voriconazole	50
		Labetalol	97	Cimetidine	54					**Norfloxacin**	87	Carteolol	44	**Thiabendazole**	47
		Phenytoin	94	Esmolol	51					Oxtriphylline	84	Hydrocortisone	41	Quinidine barbiturate	47
		Metoprolol	94	Carteolol	51					Dicumarol	84	Bevantolol	41	Pentobarbital	47
		Digoxin	94	Carvedilol	51					Theophylline	84	Paramethasone	41	Bevantolol	45
		Pindolol	94	Phenmetrazine	49					Mexiletine	84	Lithium	38	Butalbital	45
		Acebutolol	91	Dopamine	49					**Grepafloxacin**	82	Disopyramide	38	Amobarbital	45
		Chlorpropamide	91	Bisoprolol	46					Warfarin	82	Repaglinide	38	Aprobarbital	45
		Carvedilol	91	Acebutolol	43					Sotalol	82	Zuclopenthixol	38	Butethal	45
		Indomethacin	88	Dobutamine	43					Anisindione	82	Temsirolimus	38	Heptabarbital	45
		Phenobarbital	84	Metronidazole	43					Cimetidine	79			Hexobarbital	45
		Dicumarol	84	Bevantolol	40					Carteolol	79			Talbutal	45
		Atenolol	84	Practolol	40					**Ciprofloxacin**	79			Methylphenobarbital	45
		Anisindione	84	Rilpivirine	37					Ethinyl Estradiol	76			Secobarbital	45
		Fosphenytoin	81	Ofloxacin	34					Phenytoin	74			Methohexital	45
		Ibuprofen	81	Norepinephrine	31					Fluvoxamine	71			Delavirdine	45
		Penbutolol	78	Zalcitabine	31					Penbutolol	71			Cyclosporine	42
		Quinidine barbiturate	75	Sotalol	31					Prednisolone	71			Tubocurarine	42
		Gliclazide	75	Thiothixene	29					Dyphylline	68			Mivacurium	42
		Treprostinil	72	Chlorprothixene	26					Betaxolol	68			Dexamethasone	42
		Glyburide	72	Phenelzine	23					Prednisone	66			Butabarbital	42
		Cyclosporine	69							Erythromycin	66			Zafirlukast	42
		Tolterodine	69							Nicardipine	66			Doxacurium chloride	42
		Trazodone	66							Fosphenytoin	63			Ethotoin	39
		Betaxolol	66							**Terbinafine**	63			Troleandomycin	39
		Ethotoin	66							Propafenone	61			Metronidazole	39
		Glipizide	66							Clarithromycin	61			Methylprednisolone	39
		Glisoxepide	63							Ticlopidine	61			Enoxacin	39
		Tolazamide	63							Zileuton	61				

Table 5.1: Top predicted interactions for acetaminophen, pyridoxine, creatine, L-carnitine, taurine and caffeine. Interactions are ranked from highest to lowest based on the number of times they were predicted by the model. True interactions are in bold.

Acetaminophen	**Caffeine**
Acenocoumarol	Adenosine
Anisindione	Ciprofloxacin
Dicumarol	Clozapine
Imatinib	Conivaptan
Isoniazid	Grepafloxacine
Warfarin	Lithium
Pyridoxine	Norfloxacin
NONE	Regadenoson
Creatine	Tamsulosin
NONE	Ternbinafine
L-carnitine	Thiabendazole
NONE	Tolterodine
Taurine	Vemurafenib
NONE	

Table 5.2: Known (true) interactions for acetaminophen and select ingredients in energy drinks

Known interactions with acetaminophen and caffeine (see Table 5.2) were found to share common P450s, similar indications and targets. For example, many drugs that interact with acetaminophen are anticoagulants (acenocoumarol, anisindione, dicumarol and warfarin) and interact through common pathways. Others (imatinib and isoniazid), were found to likely interact through P450 metabolism and have common enzymes. Similar trends were also seen for caffeine.

Out of the 5 highest ranked top predicted interactions for acetaminophen (Table 5.1), 2 of 5 predictions are true interactions (anisindione and dicumarol). According to DrugBank [23], Timolol, like anisindione and dicumarol, is an anticoagulant. Timolol is used in the treatment of hypertension as are the 2 remaining top predicted drugs, trandolapril and treprostinil. Timolol, is primarily metabolized by CYP2D6, one of the primary P450s responsible for the metabolism of acetaminophen. Similar results are seen consistently for pryidoxine, creatine, L-carnitine, taurine and caffeine. For example, didanosin and lamivudine, predicted interactions for creatine, are both used in the treatment of the human immunodeficiency virus (HIV). Results, adapted from DrugBank [23], are summarized in Tables 5.3 - 5.8.

Acetaminophen	
Description	Analgesic and antipyretic
Targets	Prostaglandin G/H synthase inhibitor
Enzymes	CYP1A2, CYP2E1, CYP2D6, CYP1A1, CYP2A6, CYP2C8, CYP2C9, CYP3A4
Timolol	
Description	Proposed antihypertensive, antiarrhythmic, anticoagulant. Used in the treatment of migraine disorders and tremor
Targets	Beta adrenergic receptor antagonist
Enzymes	CYP2D6 and CYP2C19 substrate
Trandolapril	
Description	Prodrug belonging to angiotensin-converting enzyme inhibitor class. Metabolized in liver, regulates blood pressure and used to treat hypertension
Targets	Angiotensin-converting enzyme inhibitor
Enzymes	Liver carboxylesterase substrate
Anisindione	
Description	Anticoagulant
Targets	Vitamin K-dependent gamma carboxylase inhibitor
Enzymes	
Treprostinil	
Description	Analog of prostacyclin. Used to treat hypertension
Targets	Prostacyclin receptor agonist, peroxisome proliferator-activated receptor delta, purino receptor
Enzymes	CYP2C9 substrate
Dicumarol	
Description	Anticoagulant
Targets	Vitamin K-epoxide reductase complex subunit inhibitor, NAD(P)H dehydrpgenase inihibitor, quinone oxireductase inhibitor
Enzymes	CYP2C9 substrate, inhibitor

Table 5.3: Adapted from DrugBank. Table summarizing descriptions, targets and enzymes of top drugs predicted to interact with acetaminophen.

Pyridoxine	
Description	Coenzyme for synthesis of amino acids and neurotransmitters
Targets	Pyridoxal kinase ligand
Enzymes	CYP1A1 inhibitor
Verapamil	
Description	Calcium channel blocker. Antiarrhythmic
Targets	Calcium, potassium and sodium channels
Enzymes	CYP3A4, CYP3A5, CYP3A7, CYP1A2, CYP2C9, CYP2C8, CYP2C19, CYP2C18, CYP2B6, CYP2D6
Nadolol	
Description	Adrenergic antagonist. Used to treat arrythmias, angina pectoris and hypertension
Targets	Beta adrenergic receptors
Enzymes	
Aminophylline	
Description	Used in the treatment of lung diseases asthma, bronchitis, COPD
Targets	Phosphodiesterase inhibitor, adenosine receptor blocker, histone deacetylase activator
Enzymes	CYP1A2, CYP2E1, CYP3A4
Esmolol	
Description	Decreases force and rate of heart contractions
Targets	Beta blocker
Enzymes	
Theophylline	
Description	Smooth muscle relaxant, bronchial dilation and central nervous system stimulant
Targets	Phosphodiesterase inhibitor, adenosine receptor blocker, histone deacetylase activator
Enzymes	CYP3A4, CYP1A2, CYP2E1, CYP1A1, CYP1B1, CYP2C8, CYP2C9, CYP2D6

Table 5.4: Adapted from DrugBank. Table summarizing descriptions, targets and enzymes of top drugs predicted to interact with pyridoxine.

Creatine	
Description	Amino acid occurring in tissue and urine
Targets	Creatine kinase ligand, sodium- and chloride- dependent creatine transporter, guanidinoacetate N-methyltransferase product
Enzymes	
Labetalol	
Description	Antihypertensive
Targets	Alpha and beta adrenergic receptors
Enzymes	CYP2D6 substrate and inhibitor
Pargyline	
Description	Antihypertensive
Targets	Amine oxidase inhibitor
Enzymes	
Didanosine	
Description	Dideoxynucleoside compound. HIV replication inhibitor
Targets	reverse transcriptase inhibitor, purine nucleoside phosphorylase
Enzymes	
Lamivudine	
Description	Used in treatment of HIV and hepatitis B
Targets	Reverse transcriptase inhibitor
Enzymes	Deoxycytidine kinase, UMP-CMP kinase, phosphoglycerate kinase, nucleoside diphosphate kinase, choline-phosphate cytidylyltransferase, ethanolamine-phosphate, 5'(3')-deoxyribonucleotidase
Glucosamine	
Description	Used in the treatment of osteoarthritis. Amino sugar and precursor in biochemical synthesis of glycosylated proteins and lipids. Used to rebuild cartilage and treat arthritis
Targets	Matrix metalloproteinase antagonist, nuclear factor subunit antagonist, tumor necrosis factor, interferon gamma, chitosanase
Enzymes	CYP2C19, CYP2E1 substrates

Table 5.5: Adapted from DrugBank. Table summarizing descriptions, targets and enzymes of top drugs predicted to interact with creatine.

L-carnitine	
Description	Constituent of striated muscle and liver. Used to stimulate gastric and pancreatic secretions and treatment of hyperlipoproteinemias
Targets	Solute carrier family members, carnitine acetyltransferase, mitochondrial carnitine carriers proteins, xanthine dehydrogenase/oxidase, liver carboxylesterase, myeloperoxidase
Enzymes	
Betaxolol	
Description	Cardioselective beta adrenergic antagonist
Targets	Beta adrenergic receptors
Enzymes	CYP1A2, CYP2D6
Clarithromycin	
Description	Antibiotic
Targets	Ribosomal protein inhibitor
Enzymes	CYP3A4, CYP3A5, CYP3A7, CYP2C19, CYP1A2

Table 5.6: Adapted from DrugBank. Table summarizing descriptions, targets and enzymes of top drugs predicted to interact with L-carnitine.

Taurine	
Description	Essential nutrient for development
Targets	Alpha-ketoglutarate-dependent taurine dioxygenase, choloylglycine hydrolase
Enzymes	
Montelukast	
Description	Used in the treatment of asthma and seasonal allergies
Targets	Leukotriene receptor antagonist
Enzymes	CYP2C8, CYP3A4, CYP2C9, CYP2A6, prostaglandin G/H synthase substrate
Thioguanine	
Description	Antinoplastic, antimetabolite action. Treats leukemia
Targets	DNA
Enzymes	Hypoxanthine-guanine phosphoribosyltransferase substrate
Labetalol	
Description	Antihypertensive
Targets	Alpha and beta adrenergic receptors
Enzymes	CYP2D6 substrate and inhibitor
Azilsartan medoxomil	
Description	Used in to treat hypertension
Targets	Angiotensin receptor antagonist
Enzymes	
Torasemide	
Description	Used in the management of edema associated with congestive heart failure. Used in low doses to treat hypertension
Targets	Solute carrier family inhibitor
Enzymes	CYP2C8, CYP2C9, CYP2C19

Table 5.7: Adapted from DrugBank. Table summarizing descriptions, targets and enzymes of top drugs predicted to interact with taurine.

Caffeine	
Description	Central nervous system stimulant. Relaxes smooth muscle, stimulates cardiac muscle, stimulates diuresis and treats headache
Targets	Adenosine receptor antagonist, CAMP-specific 3',5'- cyclic phosphodiesterase inhibitor, ryanodine receptor, cyclic nucleotide phosphodiesterase competitive inhibitor, DNA-dependent protein kinase
Enzymes	CYP3A5, CYP1A2, CYP3A7, CYP3A4, CYP2E1, CYP2C8, CYP2C9, CYP1A1, CYP1B1, CYP2D6
Lithium	
Description	Used to treat gout and bipolar disorder. Mood stabilizer, used to augment antidepressants, preventative for migraine disease and cluster headaches
Targets	Glycogen synthase kinase inhibitor, inositol monophosphatase inhibitor, glutamate receptor potentiator
Enzymes	
Disulfiram	
Description	Alcohol deterrent. Alters metabolism of alcohol
Targets	aldehyde dehydrogenase mitochondrial inhibitor, dopamine beta-hydroxylase inhibitor
Enzymes	CYP2E1, CYP2C9, CYP3A4, CYP3A5
Isoniazid	
Description	Antibacterial agent, used in the treatment of tuberculosis
Targets	Catalase-peroxidase, enoyl-[acyl-carrier-protein] reductase [NADH]
Enzymes	CYP3A4, CYP2E1, CYP2C9, CYP1A2, CYP2D6, CYP2C19, CYP2A6, CYP2C8
Adenosine	
Description	Used to treat supraventricular tachycardia. Antiarrhythmic
Targets	Adenosine receptor agonist
Enzymes	Adenosine deaminase, adenosine kinase
Acyclovir	
Description	Guanasine analog antiviral drug. Acts as an antimetabolite. Used to treat herpes, chickenpox and shingles
Targets	Thymine kinase potentiator, DNA polymerase catalytic subunit inhibitor
Enzymes	

Table 5.8: Adapted from DrugBank. Table summarizing descriptions, targets and enzymes of top drugs predicted to interact with caffeine.

5.1.2 K-means and DDI network

K-means clustering

Using a k value of 2 was found to separate the data the most evenly, with cluster 0 containing 3,971 drugs, and cluster 1 containing 2,561 drugs. Using a k larger than 2 resulted in unbalanced clustering; a cluster containing the majority of the drugs in the data set as well as the majority of the interacting drugs, consistently appeared for k>2 (refer to Tables 5.9 - 5.18). Cluster information and cluster scores for k=2 is shown in Tables 5.9 and 5.10 respectively.

Cluster	0	1
Number of interactions	10191	10180
Number of drugs	3971	2561
Interaction-drug ratio	2	3

Table 5.9: The number of interactions, as well as the number of drugs found in each cluster is shown for k=2. The interaction-drug ratio is calculated as the average number of drug interactions a drug has in each cluster.

Cluster	0	1
0	57	43
1	43	57

Table 5.10: Cluster scores for k=2. Scores are calculated for each drug in a cluster as the distribution of interacting drugs among all the clusters.

Cluster	0	1	2
Number of interactions	5096	2126	13149
Number of drugs	2724	503	3305
Interaction-drug ratio	1	4	3

Table 5.11: The number of interactions, as well as the number of drugs found in each cluster is shown for k=3. The interaction-drug ratio is calculated as the average number of drug interactions a drug has in each cluster.

Cluster	0	1	2
0	28	6	66
1	14	18	68
2	25	11	64

Table 5.12: Cluster scores for k=3. Scores are calculated for each drug in a cluster as the distribution of interacting drugs among all the clusters.

Cluster	0	1	2	3	4
Number of interactions	1029	843	2405	4492	11602
Number of drugs	115	561	382	2521	2953
Interaction-drug ratio	8	1	6	1	3

Table 5.13: The number of interactions, as well as the number of drugs found in each cluster is shown for k=5. The interaction-drug ratio is calculated as the average number of drug interactions a drug has in each cluster.

Cluster	0	1	2	3	4
0	11	7	11	12	59
1	9	12	12	12	55
2	5	4	22	24	45
3	3	2	13	25	57
4	5	4	9	22	60

Table 5.14: Cluster scores for k=5. Scores are calculated for each drug in a cluster as the distribution of interacting drugs among all the clusters.

Cluster	0	1	2	3	4	5	6
Number of interactions	11404	570	4374	586	1417	41	1979
Number of drugs	2692	72	2154	795	392	89	338
Interaction-drug ratio	4	7	2	0	3	0	5

Table 5.15: The number of interactions, as well as the number of drugs found in each cluster is shown for k=7. The interaction-drug ratio is calculated as the average number of drug interactions a drug has in each cluster.

Cluster	0	1	2	3	4	5	6
0	59	2	22	2	8	0	7
1	50	7	12	6	18	0	6
2	57	2	23	4	3	0	11
3	33	6	27	12	9	1	11
4	63	7	9	4	7	0	9
5	63	5	10	12	5	0	5
6	41	2	25	3	6	0	22

Table 5.16: Cluster scores for k=7. Scores are calculated for each drug in a cluster as the distribution of interacting drugs among all the clusters.

Cluster	0	1	2	3	4	5	6	7	8	9
Number of interactions	1829	1196	1359	7184	25	41	991	6960	0	786
Number of drugs	336	263	1304	1831	203	89	88	1647	2	769
Interaction-drug ratio	5	4	1	3	0	0	11	4	0	1

Table 5.17: The number of interactions, as well as the number of drugs found in each cluster is shown for k=10. The interaction-drug ratio is calculated as the average number of drug interactions a drug has in each cluster.

Cluster	0	1	2	3	4	5	6	7	8	9
0	12	6	6	39	0	0	7	24	0	4
1	9	10	6	26	0	0	4	39	0	5
2	8	5	11	30	0	0	2	39	0	5
3	10	4	6	40	0	0	6	32	0	2
4	24	4	4	32	0	0	12	16	0	8
5	5	5	5	46	0	0	5	17	0	17
6	14	5	3	41	0	0	10	20	0	6
7	6	7	8	33	0	0	3	49	0	4
8	0	0	0	0	0	0	0	0	0	0
9	10	8	8	19	0	1	8	33	0	12

Table 5.18: Cluster scores for k=10. Scores are calculated for each drug in a cluster as the distribution of interacting drugs among all the clusters.

DDI network

From Figure 5.3 on the next page (non-connected drugs are not shown), it can be seen that the DDI network constructed from known interactions is compact and highly connected; drugs with many connections (interactions) are held together more tightly. Overlaying cluster assignments onto the network revealed clustering near the center of the network. Additionally, when incorporating P450 information, it was found that the majority (62%) of the P450 labels fall into cluster 0, with the exception of CYP3A4 which almost entirely falls into cluster 1 (Table 5.19 on page 58).

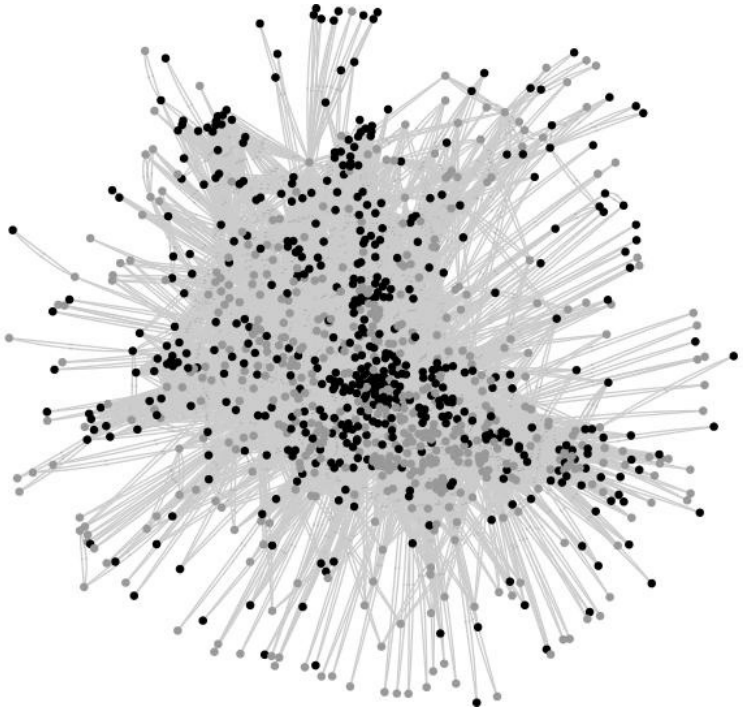


Figure 5.3: DDI network constructed from known interactions. Cluster assignments for k=2 were overlaid onto the network. Assignments for cluster 0 are in grey and are black for cluster 1.

Cluster 0		
P450	%Cluster (%)	%P450 (%)
1A2	15	86
2B6	7	85
2C8	3	54
2C9	12	61
2C19	16	75
2D6	27	73
2E1	6	100
3A4,5,7	14	29
Cluster 1		
1A2	3	14
2B6	2	15
2C8	4	46
2C9	12	39
2C19	9	25
2D6	16	27
2E1	0	0
3A4,5,7	55	71

Table 5.19: The distribution of known P450 information among clusters for k=2. %Cluster is the proportion of all the P450s in the cluster occupied by each individual P450. %P450 is the proportion of the individual P450 that is in the cluster. It can be seen from %P450 that the majority of drugs metabolized by known P450s are in cluster 0.

5.2 *In vitro* model validation

5.2.1 Transfection

Successful transfection of Hepa1-6 with CYP2E1 was not seen. Although several variables were manipulated in order to achieve a successful transfection, plasmid DNA degradation may have been one of many contributing factors.

5.2.2 Cytotoxicity

There was no significant change in cell death observed between acetaminophen and pyridoxine, creatine, L-carnitine, taurine or caffeine ($p<0.05$) when compared with acetaminophen alone. There was also no significant change in cell death observed between acetaminophen and pyridoxine, creatine, L-carnitine or caffeine ($p<0.05$) when compared with each compound alone. Further investigation is required in order to asses the possible DDI found between acetaminophen and taurine. *In vitro* evaluation of interactions with theophylline are ongoing. A summary of the results of the LDH assay are shown in Table 5.20 on the next page.

Treatment	**Average Percent Viability (%)**
Acetaminophen	94
Acetaminophen and Taurine	95
Acetaminophen and Caffeine	92
Acetaminophen and L-carnitine	96
Acetaminophen and Creatine	96
Acetaminophen and Pyridoxine	94
Taurine	93
Caffeine	92
L-carnitine	93
Creatine	95
Pyridoxine	95
Cells (no treatment)	91

Table 5.20: Summary of the results of the LDH assay. Cell death is represented as average percent cell viability (Equation 4.2). Each entry represents the mean of duplicate values of three independent experiments (n=3).

Chapter 6

Conclusions and Future Work

6.1 Summary and discussion

6.1.1 SVM

Building a classifier on a data set containing 6,532 drugs, with 42,646,653 non-interacting drug pairs and only 20,371 drug pairs known to interact (less than 0.05% of possible interactions are known DDIs), resulted in a model that is biased to predicting non-interacting drug pairs better than interacting drug pairs (99% and 49% respectively). It is important to consider that drug pairs may be labeled as non-interacting simply because their interaction is not known. This both increases the bias in the model, and also hinders its performance since the model is making predictions based on potentially false information. Using a classifier that assigns weights to the class labels so that non-interacting drug pairs and interacting drug pairs are represented evenly, improved the performance of the model, increasing the proportion of correctly predicted interactions from 25% to 50% (Table 5.2 on page 42). Nevertheless, the model was able to predict almost half of known DDIs, an event that occurs less than 0.05% of the time.

An important limitation of evaluating the model's performance using a confusion matrix is that predicted DDIs that are not previously known are considered as false positives and lower the positive predictive value of the model. Further *in vitro* validation would be needed to determine if these predictions are actual false positives, or in fact new, unknown interactions.

There were no interactions predicted between acetaminophen and pyridoxine, creatine, L-carnitine, taurine or caffeine. This was supported by *in vitro* validation and is discussed further in Section 6.1.3 on the next page. True interactions for acetaminophen and caffeine were found to occur through a variety of mechanisms; using DrugBank [23] it was found that drugs that interact with acetaminophen or caffeine share common P450s, targets and indications. The successful prediction of these interactions by the model indicate that the model is not limited to predicting interactions that occur through only one mechanism, and encompass the complex nature of DDIs.

Interactions predicted for acetaminophen, pyridoxine, creatine, L-carniting, taurine and caffeine were also found (using DrugBank [23]) to share common indications, targets and enzymes. For example, timolol, a predicted interaction for acetaminophen that is not currently known, was found to be metabolised primarily by CYP2D6. One of the main metabolizing enzymes of acetaminophen is also CYP2D6, making the prediction a valid and potential P450 mediated interaction. These trends, seen consistently among the data, support the likelihood of the predictions, however further *in vitro* analysis would be required for validation.

6.1.2 K-means and DDI network

K-means was not found to be a good separator of the data above k=2. This may indicate that K-means is not ideal in finding patterns in the data that might correlate structural features to a DDI. This is perhaps due to the simplicity of the algorithm itself; K-means may not be well suited to capturing the complexity of DDIs. Incorporating P450 information did however reveal that K-means can be used to separate drugs metabolized by CYP3A4 from other P450s, and can also separate drugs metabolized by P450s from drugs metabolized by other enzymes. This is seen in Table 5.19, where the majority of P450 metabolized drugs are localized in one cluster with the exception of drugs metabolized by CYP3A4.

The network constructed from known DDIs was found to be highly connected and compact. Because drugs with many connections are held together more tightly, the compactness of the network shows that drugs with many interactions are also connected to other drugs with many connections. Overlaying cluster assignments on the network reveals that there exists a group of drugs that, due to underlying structural factors, have a high number of DDIs. This clustering is seen in the center of the network, where drugs with the most connections are located.

6.1.3 *In vitro* model validation

Analysis of the results of the LDH assay did not reveal any interactions between acetaminophen and pyridoxine, creatine, L-carnitine, taurine or caffeine. This supports predictions made by the QSAR model, however *in vitro* assessment was limited due to the unsuccessful transfection of Hepa1-6 with CYP2E1. Proceeding with Hepa 1-6 cells without CYP2E1 would have detected DDIs mediated by other mechanisms if

they were to occur. Additional validation is ongoing.

6.2 Future work

6.2.1 Validating predictions

Testing the interactions predicted by the QSAR model would be an important step in evaluating the validity of the model. Specifically, choosing assays according to the potential mechanisms through which the top predicted drugs interact would provide useful evaluation and a means for model improvement.

6.2.2 Additional information for network analysis

Network analysis using clustering and P450 information was limited to the relatively small number of drugs metabolized by P450s available. More information (metabolism, enzymes, targets and indications) would be needed to analyze cluster assignments further and tools that bring these resources together in one environment may be beneficial in this regard.

6.2.3 Indirect interactions

Acetaminophen and caffeine are known to interact [7], although this interaction was not known to the model, nor was it predicted by the model. However, the QSAR model did predict several interactions that are shared between caffeine and acetaminophen, suggesting perhaps the possibility of an indirect interaction (Figure 6.1 on the next page). Further analysis of these kinds of interactions would be beneficial in evaluating the model further; investigating new interactions that may arise and determining whether or not they are true DDIs.

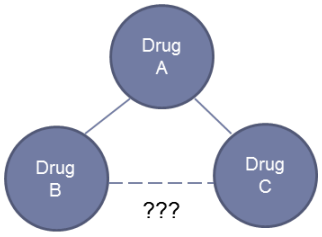


Figure 6.1: A depiction of indirect interactions. If drug A interacts with drugs B and C, how likely is it that drugs B and C also interact?

6.3 Conclusion

The QSAR Model did not predict any interactions between acetaminophen and ingredients found in energy drinks which was supported by *in vitro* validation. Although various aspects of the model have certain limitations, given the scope and feasibility of the study, a method that predicts almost half of known DDIs that are mediated through different mechanisms, was successfully developed. The model also addressed some of the other limitations of current approaches; it does not rely on *in vitro* information to make predictions and the results (the predicted DDIs) are interpretable. Further validation of the predictions is ongoing. Additional information and tools that bring more information together would assist with further interpretability of predicted interactions as well as further validation. The QSAR model developed may nevertheless provide a good basis during *in vitro* testing of DDIs. The model may not only expedite drug discovery and development, but also help identify potential DDIs between drugs that are currently on the market.

Bibliography

[1] R. H. Andres, A. D. Ducray, A. Pérez-Bouza, U. Schlattner, A. W. Huber, S. H. Krebs, R. W. Seiler, T. Wallimann, and H. R. Widmer. Creatine supplementation improves dopaminergic cell survival and protects against MPP+ toxicity in an organotype tissue culture system. *Cell Transplant*, 14:537–550, 2005.

[2] D. Arthur and S. Vassilvitskii. How slow is the k-means method? In *Proceedings of the Twenty-second Annual Symposium on Computational Geometry*, SCG '06, pages 144–153, New York, NY, USA, 2006. ACM.

[3] Z. Bibi. Role of cytochrome P450 in drug interactions. *Nutrition & Metabolism*, 5, 2008.

[4] D. K. Blumenthal and J. C. Garrison. Chapter 3. Pharmacodynamics: Molecular mechanisms of drug action. In L. L. Brunton, B. A. Chabner, and B. C. Knollmann, editors, *Goodman & Gilman's The Pharmacological Basis of Therapeutics, 12e.* The McGraw-Hill Companies, New York, NY, 2011.

[5] M. A. Correia. Chapter 4. Drug biotransformation. In B. G. Katzung, S. B. Masters, and A. J. Trevor, editors, *Basic & Clinical Pharmacology, 12e.* The McGraw-Hill Companies, Inc., New York, NY, 2012.

[6] M. J. Cupp and T. S. Tracy. Cytochrome P450: New nomenclature and clinical implications. *American Family Physician*, 57:107–116, 1998.

[7] K. DiPetrillo, S. Wood, V. Kostrubsky, K. Chatfield, J. Bement, S. Wrighton, E. Jeffrey, P. Sinclair, and J. Sinclair. Effect of caffeine on acetaminophen hepatotoxicity in cultured hepatocytes treated with ethanol and isopentanol. *Toxicology and Applied Pharmacology*, 185:91–97, 2002.

[8] J. D. Duke, X. Han, Z. Wang, A. Subhadarshini, S. D. Karnik, X. Li, S. D. Hall, Y. Jin, J. T. Callaghan, M. J. Overhage, D. A. Flockhart, R. M. Strother, S. K. Quinney, and L. Li. Literature based drug interaction prediction with clinical assessment using electronic medical records: Novel myopathy associated drug interactions. *PLoS Computational Biology*, 8, 2012.

[9] D. N. Fish. Fluoroquinolone adverse drug effects and drug interacions. *Pharmacotherapy*, 21, 2001.

[10] T. S. Furey, N. Cristianini, N. Duffy, D. W. Bednarski, M. Schummer, and D. Haussler. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16:906–914, 2000.

[11] L. Galluzzi, S. Marsili, I. Vitale, J. Michels, E. Garcia, P. Vacchelli, E. Chatelut, M. Castedo, and G. Kroemer. Vitamin B6 metabolism influences the intracellular accumulation of cisplatin. *Cell Cycle*, 12:417–421, 2013.

[12] M. Goodarzi, B. Dejaegher, and Y. V. Heyden. Feature selection methods in QSAR studies. *Journal of AOAC International*, 95:636–650, 2012.

[13] A. Gottlieb, G. Y. Stein, Y. Oron, E. Ruppin, and R. Sharan. INDI: A computational framework for inferring drug interactions and their associated recommendations. *Molecualr Systems Biology*, 8, 2012.

[14] P. F. Guengerich. Cytochrome P450 and chemical toxicology. *Chemical Research in Toxicology*, 21:70–83, 2008.

[15] E. J. Guest, K. Rowland-Yeo, A. Rostami-Hodjegan, G. T. Tucker, J. B. Houston, and A. Galetin. Assessment of algorithms for predicting drug-drug interactions via inhibition mechanisms: Comparison of dynamic and static models. *British Journal of Clinical Pharmacology*, 71:72–87, 2010.

[16] R. Harpaz, K. Haerian, H. S. Chase, and C. Friedman. Statistical mining of potential drug interaction adverse effects in FDA's spontaneous reporting system. *AMIA Annual Symposium Proceedings Archive*, pages 281–285, 2010.

[17] J. R. Horn. Chapter 66. Important drug interactions & their mechanisms. In B. G. Katzung, S. B. Masters, and A. J. Trevor, editors, *Basic & Clinical Pharmacology*. The McGraw-Hill Companies, Inc., New York, NY, 2012.

[18] J. Huang, C. Niu, C. D. Green, L. Yang, H. Mei, and J. J. Han. Systemic prediction of pharmacodynamic drug-drug interactions through protein-protein interaction network. *PLoS Computational Biology*, 9, 2013.

[19] M. G. Hudelson, N. S. Ketkar, L. B. Holder, T. J. Carlson, C. Peng, B. J. Waldher, and J. P. Jones. High confidence predictions of drug-drug interactions: Predicting affinities for cytochrome P450 2C9 with multiple computational methods. *Journal of Medical Chemistry*, 51:648–654, 2008.

[20] I. Jolliffe. *Encyclopedia of Statistics in Behavioral Science*. John Wiley & Sons, Ltd, 2005.

[21] S. M. Kakar, M. F. Paine, P. W. Stewart, and P. B. Watkins. 6'7'-dihydroxyberamottin contributes to the grapefruit juice effect. *Clinical Pharmacology and Therapeutics*, 75:569–579, 2004.

[22] B. G. Katzung. Chapter 1. Introduction. In B. G. Katzung, S. B. Masters, and A. J. Trevor, editors, *Basic & Clinical Pharmacology, 12e*. The McGraw-Hill Companies, Inc., New York, NY, 2012.

[23] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo, and D. S. Wishart. Drugbank 3.0: A comprehensive resource for 'omics' research on drugs. *Nucleic Acids Research*, 39:1035–1041, 2011.

[24] J. Lazarou, B. H. Pomeranz, and P. N. Corey. Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. *JAMA*, 279:1200–1205, 1998.

[25] H. Li, J. Sun, X. Fan, X. Sui, L. Zhang, Y. Wang, and Z. He. Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction. *Journal of Computer-Aided Molecular Design*, 22:843–855, 2008.

[26] H. Liu and H. Motoda. *Computational Methods of Feature Selection*. Chapman & Hall/CRC, 2008.

[27] E. Lounkine, M. J. Keiser, S. Whitebread, D. Mikhailove, J. Hamon, J. L. Jenkins, P. Lavan, E. Weber, A. K. Doak, S. Côté, B. K. Shoichet, and L. Urban. Large-scale prediction and testing of drug activity on side-effect targets. *Nature*, 486:361–367, 2012.

[28] T. Lynch and A. Price. The effect of cytochrom P450 metabolism on drug response, interactions and adverse effects. *American Family Physician*, 76:391–396, 2007.

[29] P. J. McDonnell and M. R. Jacobs. Hospital admissions resulting from preventable adverse drug reactions. *Ann Pharmacother*, 36:1331–1336, 2002.

[30] T. P. Minka. Automatic choice of dimensionality for PCA. *NIPS*, pages 598–604, 2000.

[31] N. Mishra. Computational modeling of P450s for toxicity prediction. *Expert Opinion on Drug Metabolism Toxicology*, 7:1211–1231, 2011.

[32] N. K. Mishra, S. Agarwal, and Raghava G. P. Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule. *BMC Pharmacology*, 10, 2010.

[33] N. M. O'Boyle and G. R. Hutchison. Cinfony - combining open source cheminformatics toolkits behind a common interface. *Chemistry Central Journal*, 2, 2008.

[34] N. M. O'Boyle, C. Morley, and G. R. Hutchison. Pybel: A Python wrapper for the OpenBabel cheminformatics toolkit. *Chemistry Central Journal*, 2, 2008.

[35] R. L. O'Connor-Semmes, K. Kersey, D. H. Williams, R. Lam, and K. M. Koch. Effect of ranitidine on the pharmacokinetics of triazolamand alpha-hydroxytriazolam in both young (19-60 years) and older (61-78 years) people. *Clinical Pharmacology & Therapeutics*, 70, 2001.

[36] The Trustees of Indiana University. P450 drug interaction table. 2011.

[37] K. C. Osterhoudt and P. M. Penning. Chapter 4. Drug toxicity and poisoning. In L. L. Brunton, B. A. Chabner, and B. C. Knollmann, editors, *Goodman & Gilman's The Pharmacological Basis of Therapeutics, 12e*. The McGraw-Hill Companies, New York, NY, 2011.

[38] F. V. Pedregosa. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[39] B. Percha, Y. Garten, and R. B. Altman. Discovery and explanation of drug-drug interactions via text mining. *Biocomputing*, pages 410–421, 2012.

[40] Pharmaceutical Research and Manufacturers of America. *Drug Discovery and Development: Understanding the R&D Process*. PhRMA, 2007.

[41] G. S. Ribas, V. Manfredini, M. G. De Marco, R. B. Vieira, C. Y. Wayhs, C. S. Vanzin, G. B. Biancini, M. Wajner, and C. R. Vargas. Prevention by L-carnitine of DNA damage induced by propionic and L-methylmalonic acids in human peripheral leukocytes in vitro. *Mutation Research*, 702:123–128, 2010.

[42] R. S. Saito, M. E. Smoot, K. Ono, J. Ruscheinski, P. Wang, S. Lotia, A. R. Pico, G. D. Bader, and T. Ideker. A travel guide to Cytoscape plugins. *Nature Methods*, 9:1069–1076, 2012.

[43] S. Schultz, M. DeSilva, T. T. Gu, M. Qiang, and K. Whang. Effects of the analgesic acetaminophen (paracetamol) and its para-aminophenol metabolite viability of mouse-cultured cortical neurons. *Basic & Clinical Pharmacology & Toxicology*, 110:141–144, 2011.

[44] S. Sheikh-Bahaei and C. A. Hunt. Enabling clearance predictions to emerge from in silico actions of quasi-autonomous hepatocyte components. *Drug Metabolism and Disposition*, 39:1910–1920, 2011.

[45] D. Slamenová, A. Gábeloá, A. Ondrejková, L. Ruzeková, T. Farkasová, and A. Collins. Preculturing of chinese hamster V79 cells with sublethal concentration of theophylline sensitizes cells to cytotoxic effects of MNNG. *Mutation Research*, 408:7–11, 1998.

[46] O. Spjuth, V. Georgiev, L. Carlsson, J. Alvarsson, A. Berg, E. Willighagen, and M. Wikberg, J. E. Eklund. Bioclipse-R: Integrating management and visualization of life science data with statistical analysis. *Bioinformatics*, 29:286–289, 2013.

[47] N. Spolaor, E. A. Cherman, and H. D. Monard, M. C. Lee. A comparison of multi-label feature selection methods using the problem transformation approach. *Electronic Notes in Theoretical Computer Science*, 292:135–151, 2013.

[48] C. Steinbeck, C. Hoppe, S. Kuhn, M. Floris, R. Guha, and E. L. Willighagen. Recent developments of the chemistry development kit (CDK) - an open-source Java library for chemo- and bioinformatics. *Current Pharmaceutical Design*, 12:2111–2120, 2006.

[49] L. Terfloth, B. Beinfait, and J. Gasteiger. Ligand-based models for the isoform specificity of cytochrome P450 3A4, 2D6, and 2C9 substrates. *Journal of Chemical Information and Modeling*, 47:1688–1701, 2007.

[50] G. Tsoumakas and I. Katakis. Multi-label classification An overview. *International Journal of Data Warehousing & Mining*, 3:1–13, 2007.

[51] P. Vasanthanathan, O. Taboureau, C. Oostenbrink, N. P. E. Vermeulen, L. Olsen, and F. S. Jorgensen. Classification of cytochrom P450 1A2 inhibitors and non-inhibitors by machine learning techniques. *Drug Metabolism and Disposition*, 37:658–664, 2009.

[52] S. Vilar, E. Uriarte, L. Santana, N. P. Tatonetti, and C. Friedman. Detection of drug-drug interaction by modeling interaction profile fingerprints. *PLoS ONE*, 8, 2013.

[53] T. C. Westfall and D. P. Westfall. Chapter 12. Adrenergic agonists and antagonists. In L. L. Brunton, B. A. Chabner, and B. C. Knollmann, editors, *Goodman & Gilman's The Pharmacological Basis of Therapeutics, 12e*. The McGraw-Hill Companies, New York, NY, 2011.

[54] M. Whirl-Carrillo, E. M. McDonagh, J. M. Hebert, L. Gong, K. Sangkuhl, C. F. Thorn, R. B. Altman, and T. E. Klein. Pharmacogenomics knowledge for personalized medicine. *Clinical Pharmacology and Therapeutics*, 92:414–417, 2012.

[55] C. Wu, D. O. Kennedy, Y. Yano, S. Otani, and I. Matsui-Yuasa. Thiols and polyamines in the cytoprotective effect of taurine on carbon tetrachloride-induced hepatotoxicity. *Journal of Biochemical and Molecular Toxicology*, 13:71–76, 1999.

[56] M. Zastrow. Chapter 2. Drug receptors & pharmacodynamics. In B. G. Katzung, S. B. Masters, and A. J. Trevor, editors, *Basic & Clinical Pharmacology, 12e*. The McGraw-Hill Companies, Inc., New York, NY, 2012.

[57] L. Zhang, Y. D. Zhang, P. Zhao, and S. Huang. Predicting drug-drug interactions: An FDA perspective. *The AAPS Journal*, 11:300–306, 2009.

Appendix A

Additional Information

A.1 Complete list of QSAR descriptors

Constitutional Descriptors

- Element Count
- Aromatic Atoms Count
- Aromatic Bonds Count
- Bond Count
- Proton belonging to an aromatic system
- Proton belonging to a pi-system
- Lipinski's Rule of Five
- Period of an atom
- Pi-contact of two atoms
- XLogP
- Amino Acids Count

Electronic Descriptors

- Atomic Polarizabilities
- Atomic Valence
- Bond Polarizabilities
- Effective Polarizability
- Hydrogen Bond Acceptors
- Hydrogen Bond Donors
- Inductive atomic hardness
- Inductive atomic softness
- Proton Total Partial Charge
- Partial Pi Charge
- Partial Sigma Charge
- Partial Total Charge (PEOE)
- Partial Total Charge (MMFF94)
- RDF Proton Descriptor
- Sigma Electronegativity
- Pi Electronegativity
- Topological Polar Surface Area
- Bond Partial Pi Charge
- Bond Partial Sigma Charge
- Bond Partial Total Charge
- Bond Sigma Electronegativity
- Charged Partial Surface Area

Geometrical Descriptors

- Distance to Atom
- Gravitational Index
- Gravitational Index (Square and Cube Roots)
- Moments of Inertia
- RDF Proton Descriptor

Topological Descriptors

- Atomic Degree
- Bonds to Atom
- Connectivity index (order 0)
- Carbon connectivity index (order 0)
- Connectivity index (order 1)
- Carbon connectivity index (order 1)
- Kier and Hall kappa molecular shape indices
- Petitjean Number
- RDF Proton Descriptor
- Topological Polar Surface Area
- Wiener Numbers
- Vertex adjacency information magnitude
- Valence connectivity index (order 1)
- Valence carbon connectivity index (order 1)
- Valence connectivity index (order 0)
- Valence carbon connectivity index (order 0)
- Zagreb Index
- Charged Partial Surface Area
- Largest Chain
- Largest Pi System
- Largest Aliphatic Chain
- Petitjean Shape Indices
- Weighted Path

Molecular Descriptors

- Atomic Hybridization
- Atomic Hybridization VSEPR
- BCUT
- Gravitational Index
- Rotatable Bonds Count
- Molecular Weight
- WHIM
- Van der Waals Radius
- Covalent Radius