Computational Estimation of Biliary Excretion of Compounds and the Role of Transporters

Thesis By

Mohsen Sharifi

In Partial Fulfillment of the Requirement

for the Degree of

Doctor of Philosophy

University of Kent

Acknowledgements

I would like to express the deepest appreciation to my principal supervisor, who was so patient with me. Dr Tara Ghafourian, who embodied the figure of a kind mentor, provided support and instilled mental stimulation in regards to my research. I would never have been able to finish my dissertation without her guidance. Dr Ghafourian inspirerd me to take pride in my research. She patiently corrected my writing, her enthusiasm for research efforts will have a significant effect on my future projects. Providing support letters, she helps me to win a number of student bursaries to attend different conferences and workshops. Without her persistence, I would have lost my motivation. Dr Ghafourian gave direction when it was most needed. Her valuable advice throughout this process kept me focused on my research topic. I say a big THANK YOU!

Special thanks goes to my second supervisor, Dr Ali Nokhodchi for his brilliant suggestions and supports.

I would like to express my special appreciation and thanks to Prof. Iain Cumming for all his helps and supports. Also I would like to thank the administration group of Medway School of Pharmacy, especially Richard Hammersley and Tom Bore for their excellent guidance. I also thank the wonderful staff in the Chemistry Department as well as in other departments for always being so helpful and friendly, particularly, Dr Vadim Sumbayev and Dr Maxwell Casely-Hayford who are genuinely nice and want to help out and I'm glad to have interacted with them. If I have forgotten anyone apologize.

Also, I am grateful to my close friends, Mehrdad Mirzaee, Ehab Al-Moubarak and Rafał Wyszyński, who helped me tremendously. Thank you guys, you always listened to my complaints and made my boring and hard life in Medway easier. Also, worthy of appreciation is to Danielle Newby for her assistance, apart from being a reliable friend, she always willing to help and gives her best suggestions. Danielle corrects my English and teaches me some interesting English idiomatic expressions such as "give someone a wedgie" and "going to getting smashed".

Besides, I heartily thankful to Prof. Martin Michaelis at School of Biosciences, Kent University for giving me the opportunity to participate in collaborative projects on anticancer drugs.

I gladly express my gratitude to University security staffs at the Medway's Gate House, they are really helpful and let me work till very late and on all Saturdays and Sundays.

Lastly, I would like to thank my family for all their love and encouragement, especially my dad, Dr Sharifi, who fully supported me financially for my PhD and also, always guided me in my life. Also, I would like to thank my mom who support me morally and prayed for me. Without them I wouldn't be where I am today.

Table of Contents

List of Tables

List of Figures

Abstract

Biliary excretion is one of the main elimination pathways for drugs and/or their metabolites. Therefore, an insight into the structural profile of cholephilic compounds through accurate modelling of the biliary excretion is important for the estimation of clinical pharmacokinetics in early stages of drug discovery. The aim of this project was to develop Quantitative Structure-Activity Relationships (QSAR) as computational tools for the estimation of biliary excretion. In addition, the structural requirements for biliary excretion were investigated in relation to the structural requirements for binding to uptake and efflux transporter proteins that are involved in hepatobiliary elimination.

The study used three datasets; 1. percentage of dose excreted intact into bile in rat for 217 compounds, 2. P-gp inhibition constants for 219 compound, 3. percentage inhibition of OATP transporters, OATP1B1, OATP1B3 and OATP2B1. Statistical techniques were stepwise regression analysis, Classification and Regression Trees (C&RT), Chi-square Automatic Interaction Detector (CHAID), Boosted trees (BT), Random Forest (RF) and Multivariate Adaptive Regression Splines (MARS) models.

The study resulted in QSARs for the prediction of biliary excretion, P-gp binding constants and percentage inhibition of OATPs, along with QSARs incorporating predicted P-gp and OATP inhibition values for the prediction of biliary excretion. Simple regression tree models were of similar accuracy to the boosted trees model in the estimation of the percentage of bile excretion of compounds. Molecular descriptors selected by these models indicated a higher biliary excretion for relatively hydrophilic compounds especially if they have acid/base dissociation, and a large molecular size above 348 Da.

The major role of OATPs in biliary excretion was indicated using interactive decision tree models with OATP1B1 binding being the most successful predictor of biliary excretion amongst the three OATP subfamilies. In contrast, predicted P-gp binding parameters were not successful in the prediction of biliary excretion. This may be due to problems in extrapolating the *in vitro* P-gp binding data to the *in vivo* situation, or due to the difference in the chemical spaces of the P-gp and biliary excretion datasets which may lead to the compounds in biliary excretion dataset to fall outside the applicability domain of the P-gp models.

1. Introduction

1.1. Drug Discovery and Development

Discovery and development of a drug is a very expensive process (Djulbegovic *et al.*, 2014). Toxicity, poor efficacy and poor bioavailability are the main reasons for failure during discovery, development and registration of drug candidates (Gad, 2005). Early identification of poor candidates is very essential for reducing the cost and the resources spent on drug discovery and development. For most drugs, discovery and development could be a remarkably long process. For example, from initial stage to approval of Food and Drug Administration (FDA) for taxol which is a chemotherapeutic drug, was nearly 30 years (Rowinsky *et al.*, 1990). There has been a steady decline in the number of drugs approved for marketing by regulatory agencies since the 1960s, despite the advancements in drug discovery technology, and the increasing investments of the pharmaceutical companies. The trend can be seen from 70-100 drugs introduced in the 1960s, and 60-70 drugs in the 1970s, to about 50 in the 1980s, and less than 40 in the 1990s and after (Hillisch and Hilgenfeld, 2003) (Figure 1.1).

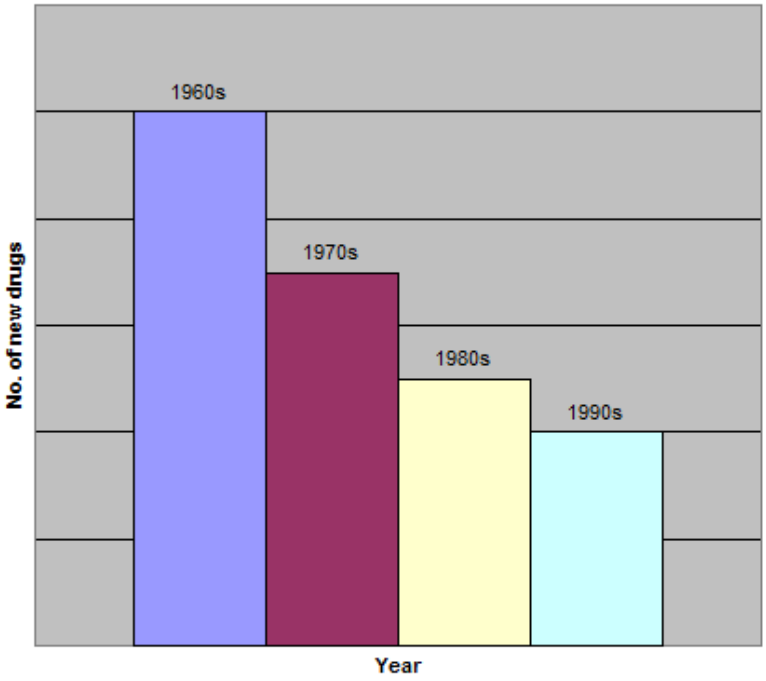

Figure 1.1. Graph showing the number of new drugs introduced from the 1960s to 1990s (Adapted from Hillisch and Hilgenfeld, 2003)

Some of the factors that are considered to be responsible for this decline are the stricter control of the process by regulatory agencies such as the FDA to ensure the safety of compounds before approval. This leads to high attrition rates and a prolonged duration of the drug development process (Hillisch and Hilgenfeld, 2003). The major cause for decline of new molecular entities (NMEs) or failures recorded in drug development was attributed to poor pharmacokinetics (39%) and animal toxicity (11%) (Waterbeemd and Gifford, 2003; Rang, 2006).

Drug candidates normally undergo prior investigation with selection of those compounds with optimal properties including physicochemical parameters (Lipinski *et al.*, 1997). According to Kerns and Di (2008) important properties in drug discovery can be classified in four groups: (1) Structural properties, e.g. hydrogen bonding, lipophilicity, molecular weight (MW), pKa, polar surface area, shape and reactivity, (2) Physicochemical properties such as solubility, permeability and chemical stability, (3) Biochemical properties, such as metabolism and transport, (4) Pharmacokinetics and toxicity, e.g. clearance, half-life, bioavailability and LD_{50}.

Initial identification of drug candidates is based mainly on the ability of compounds to have a desired activity and selectivity against a target (e.g. inhibitory effect). Investigation of other properties is traditionally postponed to later stages of the development process, due in part to the success of pharmaceutics research in achieving adequate absorption or bioavailability of drug molecules (Bleicher *et al.*, 2003). Recently, with the advent of modern technologies in drug discovery including *in silico* methods, to address the problem of high attrition rate, screening of potential drug candidates for their pharmacokinetic and physicochemical properties is being introduced by the pharmaceutical industry much earlier during drug development (Rang, 2006). A much better approach which helps facilitate the success and approval of a drug molecule is the use of predictive tools in the design phase of the synthesis of compound libraries (Waterbeemd and Gifford, 2003).

Nowadays, *in vitro* methods and statistical modeling are used extensively in the development of drugs. These methods allow the reduction in more expensive *in vivo* experiments. Model development in drug development is usually empirical or exploratory in nature. Models are developed using experimental data and then refined until a reasonable balance is obtained between overfitting and underfitting (Bonate, 2006). Computational modeling may be helpful in assay systems resulting in faster discovery of new potential drugs (Bronchud *et al.*, 2008).

The prediction ability of ADME properties as well as the knowledge of the binding/modulating properties of drug molecules on membrane transporter proteins are important as they inherently contribute to the pharmacokinetic properties. Transporters such as P-glycoproteins belong to the ATP-binding cassette superfamily of membrane transporters (Poongavanam *et al.*, 2012). The FDA has urged that every new molecular entity should be routinely checked for a possible interaction with P-glycoproteins (FDA Guidelines, 2014). Thus, in lead optimisation process, early identification of membrane transport protein ligands, being substrates or inhibitors, is of utmost importance to improve the ADME profile of drug candidates (Bleicher *et al.*, 2003; Di Pietro *et al.*, 2002).

1.2. Pharmacokinetics

Absorption, distribution, metabolism and excretion (ADME) are the main processes in biological disposition of a drug. Following drug administration, depending on the site of administration, drug concentration will increase in the blood, plasma and consequently in tissues due to the absorption process. This is followed by a decline in plasma concentration due to drug distribution into tissues and elimination. Pharmacokinetics (PK) is the study of the time course of drug concentration in the body. In addition to dosage regimen decisions, other applications of pharmacokinetics studies include bioavailability measurements, effect of physiological and pathological conditions on drug distribution, elimination and absorption, dosage adjustment of drugs in disease states when necessary, correlation of pharmacological responses with administered doses, evaluation of drug interactions and finally clinical prediction using pharmacokinetic parameters to individualize the drug dosing regimen (Jambhekar and Breen, 2009). In general, PK parameters of a drug result from its physicochemical and biochemical properties. These properties are determined by the structure of the drug (Kerns and Di, 2008).

Absorption phase is the first pharmacokinetic process before the distribution and elimination. After a standard dosage of oral administration enters the gastric fluid, the drug is gradually released from the formulation and the absorption process starts (Rosenbaum, 2011). In this phase, the dissolved drug has the chance to pass through the GI membrane into the blood. Passive absorption is thought to be the main mechanism of absorption for most drugs. However, uptake transporters (carrier proteins) in intestinal epithelial membrane may be facilitating the absorption process. Besides, in the enterocyte membrane, drug absorption may be reduced if efflux transporters take the drug back into the lumen (Rosenbaum, 2011). Absorption of proteins and macromolecular drugs from the GI tract is hard due to their large size and, therefore, parenteral administration is the predominant route of drug delivery for these drugs (Pandit, 2007). Other routes of administration include the transdermal route, when drug is applied to the skin for systemic absorption through the skin, the respiratory route, in which drug is inhaled into the lungs and the main absorption happens in the alveoli, and the nasal route, where the nasal

mucosa with a good blood supply can absorb the drugs quickly depending on the duration of drug contact with the nasal mucosa (Pandit, 2007).

Distribution is the next important phase in pharmacokinetics that controls drug concentrations in the tissues and the observed pharmacological response. Drug distribution to peripheral tissues is dependent on four main factors: (1) the drug concentration; (2) the drug physicochemical properties; (3) the blood flow to the tissue; and (4) the affinity of drug for the tissue vs. the drug affinity to plasma proteins. Amongst these factors, physicochemical properties of drugs such as acid dissociation constant and molecular weight (MW) are some of the most influential factors in tissue distribution (Riviere, 2011). Apart from the above mentioned parameters, the rate of drug metabolism plays a key role in distribution, since readily metabolised compounds are less available for tissue distribution (Riviere, 2011). Metabolism plays an essential function in the drug elimination. The rate of metabolism for drugs that are very rapidly or very slowly cleared can present problems in accurate control of the plasma levels, and, with persistent compounds of very long half-lives , the risk of toxicity can be considerable (Coleman, 2005). First-pass metabolism is a situation when a drug is metabolized prior to reaching systemic circulations. First-pass metabolism may happen in both the liver and the gut (Chesnokova *et al.*, 2007). In general, the liver is the most important and sometimes the only site of metabolism. Extensive metabolism in one or more other tissues, such as the kidney, lung and gastrointestinal membrane is rarely observed (Tozer and Rowland, 2006).

In addition to the metabolism, drug excretion by the kidneys and liver are the main routes of drug elimination. The Kidney is the main organ of excretion, while several compounds are excreted in bile. The renal excretion is mainly by glomerular filtration (Rosenbaum, 2011). Drugs that are secreted into the bile finally pass into the intestine. In the intestine they may be re-absorbed; this process is known as the enterohepatic circulation. The route and the rate of a drug's elimination has major consequences in terms of the pharmacokinetics, drug-drug interactions, and the pharmacotherapy in general. The elimination process has been discussed in a greater detail in section 1.3.

1.3. Elimination of Drugs

Drugs can be eliminated by metabolism or excretion. Excretion is the process that removes a drug from tissues and circulation (DiPiro *et al.*, 2010). Therefore, excretion in theory could include discharge into the urine, feces (via bile from the liver), exhaled air (via the lungs), or sweat (via the skin). However, for most drugs, the primary route of excretion is the renal excretion into the urine via the kidneys and/or the biliary excretion into the bile via the liver (Taft, 2009). Renal excretion is more common for the water-soluble molecules; hence, many polar drugs with low log P values are excreted unchanged directly into the urine. Lipophilic drugs may experience the process of tubular reabsorption and move from the urine (tubule of the nephron) into the peritubular capillaries, and consequently cannot be eliminated by renal excretion. For these drugs, hepatic clearance may be the main route of elimination. The primary purpose of hepatic metabolism is to create more hydrophilic molecules that will not be reabsorbed and, thus, can be excreted from the body in the urine or bile. Most drugs are lipophilic in nature and are eliminated by metabolism or biotransformation (Rosenbaum, 2011). Drug molecules that are larger (high molecular weight), and glucuronide and glutathione conjugates are more likely to be excreted via the liver into the bile. Compounds that are excreted into the bile end up in the intestines, where they may be eliminated by the feces or reabsorbed (Taft, 2009).

Clearance is a parameter that indicates the rate at which a drug is cleared from the body. It is defined as the volume of plasma from which all drug is removed in a given time presented in volume per time units (Stringer, 2006). This powerful parameter is used in pharmacokinetics for the evaluation of the elimination, and for clinical applications. Clearance may be viewed as a factor of drug elimination rate (eq. 1.1):

$$\textit{Rate of elimination} = Cl \,.\, C \qquad \textbf{Eq. 1.1}$$

Where C is the blood concentration (Tozer and Rowland, 2006). As we can see in Eq.1.1, clearance relates the rate of drug elimination to the concentration. Total clearance (Cl) or total body clearance which is referred to as systemic clearance, is

sum of all the component clearances by different body organs (Rosenbaum, 2011) given by eq. 1.2.

$$Cl = Cl_R + Cl_H + Cl_{other} \qquad \textbf{Eq. 1.2}$$

In eq. 1.2, Cl is the total body clearance, Cl_R is the renal clearance, Cl_H is the hepatic clearance, and Cl_{other} indicates any other form of clearance.

The compartmental models below show how we can calculate elimination in the body (Patric, 2006):

$$\frac{dX}{dt} = -k_{el} \,.\, X \qquad \textbf{Eq. 1.3}$$

Where X is the amount of drug in the body and t is the time after administration of dose and k_{el} shows the elimination rate constant.

Integration from the above equation presents the next expression:

$$X = X_0 \,.\, e^{-k_{el}.t} \qquad \text{then,} \qquad log\, X = log X_0 - \frac{-k_{el}.\ t}{2.303} \qquad \textbf{Eq. 1.4}$$

Where X_0 represents the initial amount of drug in the body.

Alternatively, k_{el}, can be calculated with the help of other pharmacokinetic parameters (eq. 1.5):

$$k_{el} = Cl_T / V_d \qquad \textbf{Eq. 1.5}$$

Where V_d represents the apparent volume of distribution.

Furthermore, total clearance and volume of distribution can be calculated from the following equations:

$$Cl = \frac{-(\frac{dX}{dt})}{C} \qquad \text{and} \qquad V_d = \frac{X}{C} \qquad \textbf{Eq. 1.6}$$

Here, C stands for the drug concentration in plasma.

1.3.1. Renal Excretion

Renal excretion is a very vital process by which the products of metabolism and waste metabolites are cleared from the organism (DiPiro *et al.*, 2010). Although kidneys have several functions, maintaining the homeostasis by regulating fluid and electrolyte balance is the main function of the kidney. The kidneys are responsible for the reabsorption of water, glucose, and amino acids (Pandit, 2007). Renal elimination of drugs consists of three stages of glomerular filtration, proximal tubular secretion and distal tubular reabsorption (Stringer, 2006). As it was stated before, the water-soluble materials are excreted better from the kidney (Haschek *et al.*, 2010). Acidic or basic states of a drug and pH of the urine are important parameters in the fate of a drug in renal excretion (Haschek *et al.*, 2010). Active tubular secretion and glomerular filtration are the main pathways in renal elimination (Haschek *et al.*, 2010).

A glomerulus is a big knot consisting of capillaries and surrounded by Bowman's capsule; 120 to 150 ml of blood is filtered at the glomerular capillaries per minute. The glomerular capillaries are fenestrated and freely permeable to water, electrolytes and most plasma ingredients. The pore size in these capillaries can permit most agents and drugs with the molecular weight smaller than 67 kDa to pass through and return to plasma (Smith, 2006).

If a drug does not binds to a plasma protein (such as albumin) and it is small enough to be filtered in the glomerulus, then, its clearance by glomerular filtration is equal to the glomerular filtration rate (GFR).

$$Cl_{GF} = GFR \qquad \textbf{Eq. 1.7}$$

In Eq. 1.7, Cl_{GF} is the clearance by glomerular filtration. However, many drugs bind to the plasma proteins, and bound drug will not be filtered. Here, *fu* is the unbound fraction of drug.

As a result the glomerular clearance can be calculated by Eq. 1.8 below (Janku, 1993).

$$Cl_{GF} = fu \,.\, GFR \qquad \textbf{Eq. 1.8}$$

Some of the chemicals that are filtered at the glomerulus are reabsorbed by active transport system found primarily in the proximal tubules. In proximal renal tubules, there are two systems primarily responsible for the active tubular secretion of drugs, one for organic anions and another for organic cations. The anionic system (OATs transporters) transports organic acids such as penicilins, indomethacin and glucuronides. The cationic system (OCTs transporters) transports organic bases such as morphine, procaine and quaternary ammonium compounds. Both active and passive transports are involved in tubular secretion process (Burckhardt and Wolff, 2000). It is worth mentioning that P-glycoprotein is present in the brush border of the renal proximal tubules, and can play a role in the active tubular secretion of exogenous substances. This pump is involved in tubular secretion of, for example digoxin, and can be inhibited by quinidine or verapamil, leading to an increase in digoxin serum concentrations (Giacomini *et al.*, 2010). Some drugs can inhibit the secretory function of tubules and renal clearance would reduce consequently. Probenecid which is also used in treating gout and hyperuricemia, is a good example of a drug that can inhibit tubular secretion of several agents such as verapamil (Piscitelli *et al.*, 2005).

Volume of plasma that is cleared from a compound in kidneys in unit time shows renal clearance (Cl_R) and can be calculate by equation 1.9 (Rosenbaum, 2011).

$$Cl_R = (C_{ur} \,.\, Q_R) / C \qquad \textbf{Eq. 1.9}$$

Where, Cl_R is the renal clearance of a compound, C_{ur} stands for drug concentration in the urine, C shows plasma concentration and Q_R is the urine flow rate (ml/min).

1.3.2. Elimination by the Liver

Liver is a major elimination organ which eliminates drugs by metabolism and biliary excretion. One of the most important functions of the liver is the formation of bile. However, the liver is generally identified with its primary role in drug metabolism.

Bile is a composition of bile acids and other components such as phospholipids, bilirubin and cholesterol that is formed in the canaliculus between adjacent hepatocytes and is actively discharged across the canalicular membrane. Many drugs are also excreted through this system in significant quantities (Taft, 2009). The resulting bile is stored in the gallbladder and released into the intestine. Once bile is released into the intestine, some metabolites and unchanged drugs continue their way of elimination through the feces. Others, mostly lipid-soluble drugs, are reabsorbed from the intestine and move to the systemic circulation (Luscombe and Nicholis, 1998). This process is known as enterohepatic circulation and it affects pharmacokinetics by keeping the plasma concentration high (Plusquellec *et al.*, 1998). Despite the possibility of reabsorption, bile plays an important role in the excretion of xenobiotics, including drugs and their metabolites, which is in addition to its physiologic role in the intestinal digestion of lipids and lipid-soluble vitamins. This includes a diverse array of compounds, both polar and lipophilic, including anions, cations, and neutral molecular (Taft, 2009). Elimination of some drugs, e.g. oestrogens, is very slow while water-soluble drugs are excreted in faeces through the intestine quickly (Smith, 2006). Enterohepatic cycling and biliary elimination can continue until the compound is ultimately eliminated from the body by faecal or renal excretion or metabolism.

Hepatic clearance (Cl_H) (by metabolism and/or biliary excretion) is defined as the volume of blood from which drug is removed completely by the liver per unit time. Hepatic clearance is a function of hepatic blood flow (Q_H) and the extraction efficiency of the liver for the drug (E_H) (Tozer and Rowland, 2006).

$$Cl_H = Q_H \,.\, E_H \qquad \textbf{Eq. 1.10}$$

Hepatic elimination can range from 0 (when the liver is incapable of removing the drug) to 100% (when the liver extracts the entire drug presented in a given pass). Moreover, Cl_H is equal to systemic clearance only when the drug is cleared completely by the liver after intravenous administration (Burton *et al.*, 2006).

The amount of circulating drug presented to the liver enzymes and cleared from the blood depends on the rate of hepatic blood flow (Q_H), binding to the circulating

proteins, and the metabolic activity and bile excretion involved in the hepatic elimination of the compound (Nassar *et al.*, 2009). The hepatic intrinsic clearance of unbound drug in the liver ($Cl_{u.int}$) indicates the maximal ability of hepatocyte to remove drug from the liver. In most cases $Cl_{u.int}$ will exceed the hepatic clearance of the total drug (see equation below). The hepatic intrinsic clearance of unbound drugs is frequently related to metabolic activity, which often is assumed to be the rate-limiting step in hepatic elimination:

$$Cl_{u.int} = \sum \frac{Vmax}{Km+Cu} \qquad \textbf{Eq. 1.11}$$

Where *Vmax* is the maximum rate of the reaction for enzyme involved in the metabolism of the substrate, *Km* is the concentration at which the metabolic rate will be half in the enzyme reaction and C_u is the concentration of unbound drug at the enzyme site in the liver (Burton *et al.*, 2006).

1.3.3. Elimination by the Other Sites

Beside the major routes of excretion (bile and kidney), excretion can also take place through other excretion routes such as lungs, saliva, sweat, feces, mother's milk and hair. Lungs have the main role in pulmonary excretion of some xenobiotics which exist in gaseous phase in the blood (Haschek *et al.*, 2010).

In breastfeeding mothers, unchanged drugs, drug metabolites and toxicants can be excreted into the milk as an excretion route. As milk's pH is slightly acidic at about 6.5, basic compounds are more excreted into the milk than acidic compounds.

In case of extensive sweating, study of elimination through sweat could be essential. Iron, cadmium, zinc and some other metals could be excreted in sweat (Hale *et al.*, 2002).

Feces can be the main route of elimination for any drug which is not absorbed in the small intestine or via enterohepatic circulation.

1.4. Function of the Liver and its Role in Drug Elimination

Liver is the largest internal organ in the body; it is relatively larger in infancy, comprising one-eighteenth of the birth weight (Sherlock and dooley, 2008). The liver is divided into the right and left lobes but most of the liver's mass is found in the right lobe. Anatomically this exocrine and endocrine organ is situated in a very strategical place, between pancreas, gastrointestinal tract and spleen (Figure 1.2). The entire surface of the liver is covered by a capsule that contains nerves which can sense pain (Sherlock and dooley, 2008). The gallbladder is located under the liver (Figure 1.2). The liver has a double blood supply; portal vein brings venous blood from the intestine and spleen, and the hepatic artery, coming from the celiac axis, supplies the liver with arterial blood. The liver receives approximately 1100 ml/minute of blood from the portal vein and 350 ml/minute of blood from hepatic artery (Taft, 2009).

Liver acts as a detoxifier to protect the general blood circulation from toxins that are absorbed through gastrointestinal tract. This is done through metabolism and excretion through bile. Moreover, liver is responsible for maintaining adequate blood sugar concentrations. Blood from pancreas, which is rich in glucagon and hormones, and the blood from spleen, which contains the metabolites from the red blood cell breakdown, pass through the liver via the portal vein for detoxification. Apart from the production of bile and metabolism, hepatocytes play other important functions such as destroying bacteria by the use of peroxisomes and lysosomes. A hepatocyte can contain 800-1000 mitochondria per cell. Besides, hepatocytes have many rough and smooth endoplasmic reticulums. Smooth endoplasmic reticulum produces lipids, and catabolise estrogen, progesterone and testosterone. Rough endoplasmic reticulum may synthesise plasma proteins such as albumin from amino acids and then return them back to the space of disse (You and Morris, 2007). Other functions of hepatocytes are synthesise of the alpha and beta globulin, plasma proteins, coagulation factor, very low density lipoprotein (VLDL), low density lipoprotein (LDL) and high density lipoprotein (HDL). Activation of vitamin D is another essential function of hepatocytes (Pocock and Richards, 2009).

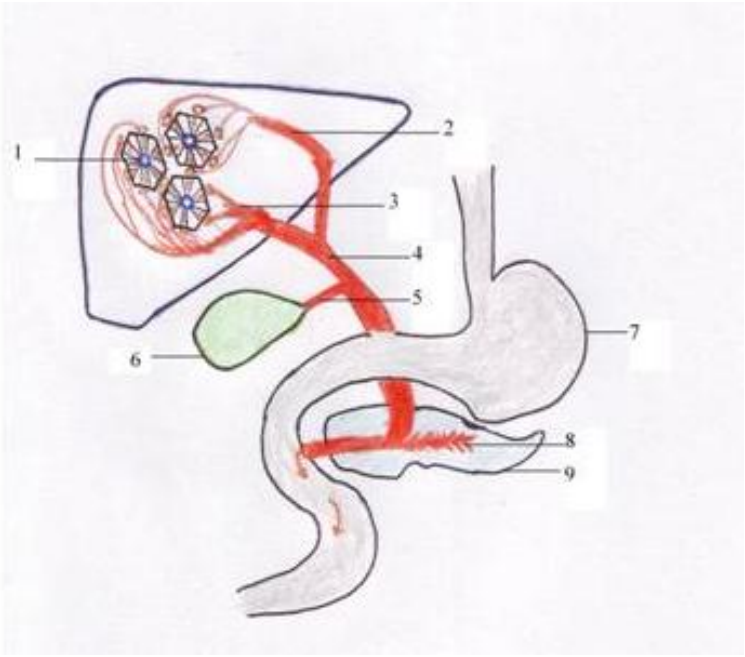

Figure 1.2. Bile release into the duodenum: 1. Hepatic lobule, 2. Left hepatic duct, 3. Right hepatic duct, 4. Common hepatic duct, 5. Cystic duct, 6. Gall bladder, 7. Stomach, 8. Pancreatic duct, 9. Pancreas (adapted from Guyton and Hall, 2006).

1.4.1. Biliary Excretion of Drugs

Functional unit of the liver is known as lobule. Figure 1.3 shows the structure of the liver‘s lobule. A lobule is defined at the histological scale and involves branches of the portal vein and hepatic artery, and a central vein in terms of the blood flow. The blood from branches of the portal vein and hepatic artery vessels eventually mix at sinusoid. In the sinusoids the mixed blood will keep moving from periphery to the centre of the lobule. The plasma near hepatocytes leaks in the area close to hepatocyte cells, which is called space of disse. All the plasma is well exposed to hepatocyte and therefore hepatocyte can efficiently exchange chemicals with plasma in the space of disse. For example, the toxin in the plasma can be detoxified or the extra glucose can be converted to glycogen by the hepatocyte and then returned to the space of disse. Central vein is situated in the centre of each lobule; the blood from portal vein and hepatic artery passes to the central vein through the sinusoid (Guyton and Hall, 2006).

Hepatocyte's one face is to the blood (via space of disse) and the other face is to the other hepatocytes. This means that, hepatocytes are laid back to back and the bile is secreted by the hepatocytes in the space between them, bile canaliculus, and then the bile duct. Excreted bile, unlike the blood flow, moves away from the centre of the lobule to the periphery (Figure 1.3). The resulting bile drains into branches of intrahepatic bile ductules that converge to the common hepatic bile duct (Matsumoto and Nakamura, 1992). Finally, the secreted bile from the left hepatic duct together with the right hepatic duct join together to make common hepatic duct. The common hepatic duct joins to the gallbladder through the cystic duct. In the healthy man, gallbladder stores about 50 ml of bile and during storage bile becomes more concentrated which increases its potency and intensifies its effect on fats (Guyton and Hall, 2006). Uptake from sinusoidal blood and then secretion of bile salts across the canalicular hepatocyte membrane are the major factors controlling the rate of bile secretion.

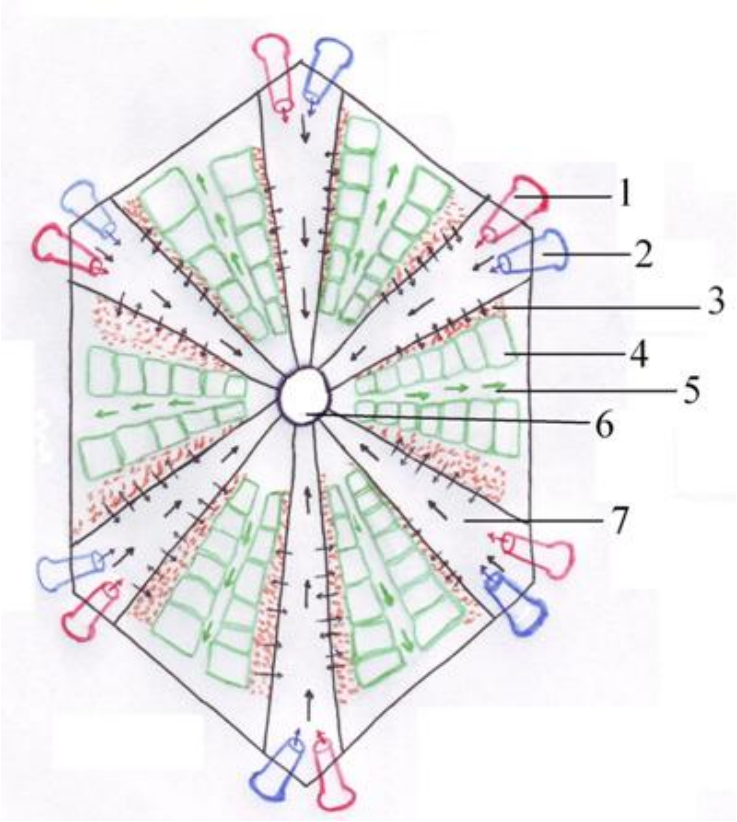

Figure 1.3. Schematic representation of bile duct and blood flow in lobule organisation. Lobule is the basic functional unit of the liver. The liver lobule is constructed around a central vein, which empties into the hepatic vein; 1. Branch of hepatic artery, 2. Branch of portal vein, 3. Space of disse, 4. Hepatocyte, 5. Bile canaliculas 6. Central vein, 7. Sinusoid (adapted from Guyton and Hall, 2006).

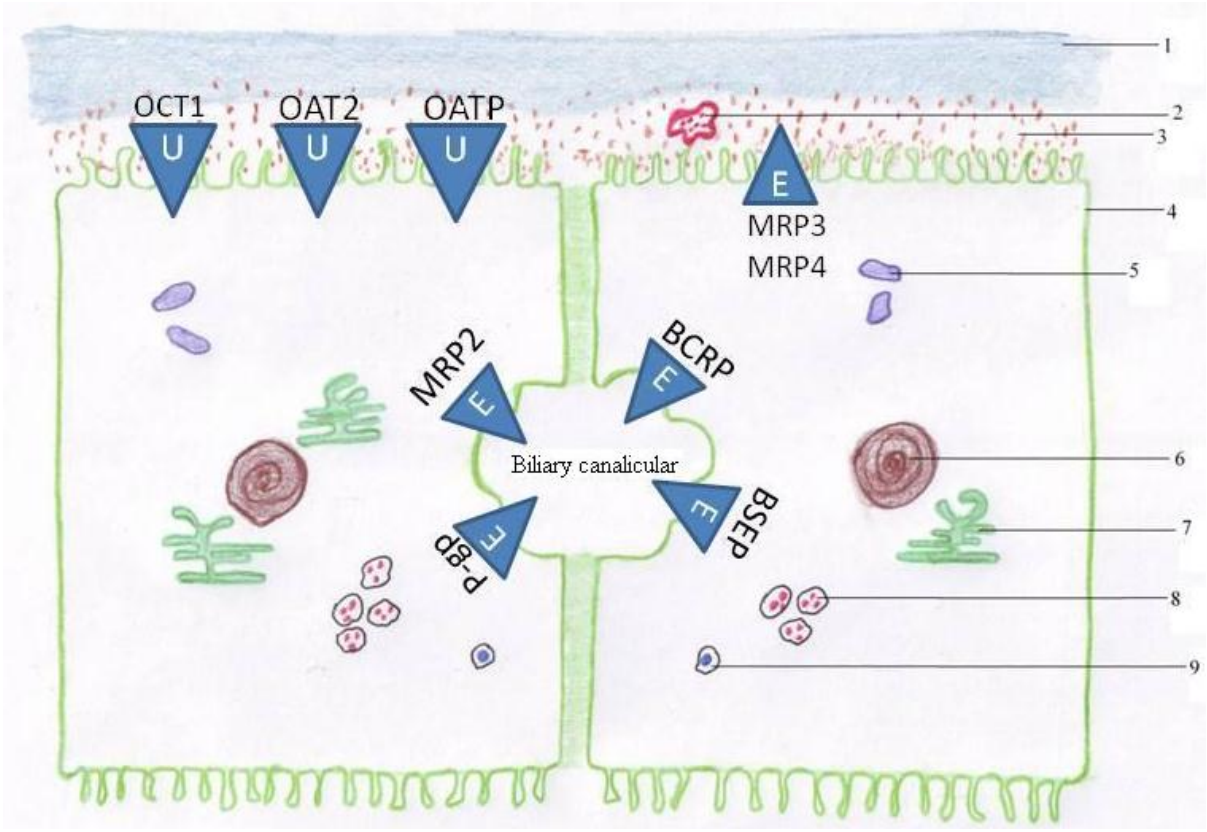

Figure. 1.4. The cartoon depicts substrate transport processes in the hepatocyte including sinusoidal and canalicular proteins efflux (E) and uptake (U) transport of drugs/drug-likes and their metabolites. 1. Sinusoidal membrane, 2. Ito cell, 3. space of Disse, 4. hepatocyte, 5. mitochondria, 6. nucleus, 7. endoplasmic reticulum, 8. lysosomes, 9. Peroxisome (Sharifi and Ghafourian, 2014)

Liver plays a very key role in drug elimination via bile. Liver is able to secrete up to 1 litre bile per day, which accumulates in gallbladder and can be emptied in duodenum for digestion of food (Pandit, 2007). The most important components of the bile are conjugated bilirubin, phospholipids and lecithin, IgA antibodies, cholesterol and bile salts such as cholic acid and chenodeoxycholic acid. Bile acids are some of the most important substances in bile that are vital for efficient digestion and emulsification of lipids. Most bile acids originate from the recirculation pool (Dawson *et al.*, 2009). Bile acids are also synthesized by the liver from cholesterol.

Canalicular bile secretion is an osmotic process in which active excretion of organic solutes into the bile canaliculus is the main driving force for the passive inflow of

water, electrolytes, and nonelectrolytes from hepatocytes (Trauner and Boyer, 2003). Several different types of transporter proteins are involved in the uptake of compounds from the blood into hepatocytes, and others are responsible for efflux of the compounds from hepatocyte into the canaliculus through canalicular membrane. These proteins are located in the basolateral and canalicular membranes of the hepatocytes and the substrate compounds include chemically diverse metabolites and unchanged drugs. Figure 1.4 shows the main transport proteins in hepatocytes that are responsible for the uptake of compounds from plasma and excretion to outside the cells. While products of the multidrug resistance gene family (MDR), namely bile salt export pumps, Bsep (rat) and BSEP (human), transport monovalent bile salts (Rollins and Klaassen, 1979), excretion of non-bile salt organic anions and divalent sulphate or glucuronide bile salts is carried mainly by the multidrug resistance protein 2 (MRP2) and P-glycoprotein. Bile salt export pump has a limited role in drug excretion (Morgan *et al.*, 2010). The transporter proteins responsible for biliary excretion have been explained in section 1.5.

Chemical structure, polarity and molecular size as well as characteristics of the liver such as specific active transport sites within the liver cell membranes are the main factors which determine elimination via the biliary tract (Rollins and Klaassen, 1979). Apart from physico-chemical factors, species, strain, gender differences and diet also can play a role in hepatic elimination. For instance, sex-dependant expression and activity of hepatic BCRP in males is higher in both mice and humans (Merino *et al.*, 2005a). Another interesting fact is that hepatic MRP2 expression in rats is nearly 10 fold higher than in humans (Li *et al.*, 2008) moreover, species differences in substrate specificities in transporters are not negligible (Takekuma *et al.*, 2007).

1.4.2. Metabolism of Drugs

The liver is the important site of metabolism for various compounds including drugs. Metabolism, or biotransformation, is a major route of elimination for many drugs. Drug metabolism often converts lipophilic compounds into more polar products. Carbohydrates, fats, and proteins are all broken down by hepatic

enzymes. A healthy liver detoxifies much of the harmful substances (Gibson and Skett, 2001), but liver diseases can affect drug metabolism and the biliary clearance (e.g., Cirrhosis, Cholestasis and Carcinoma) (Paintaud *et al.*, 1996). Studies in biliary excretion of some extensively metabolised drugs show that many patients with liver dysfunction can metabolise and excrete drugs normally, while other patients have a decreased metabolism and biliary excretion rates (Hvidberg *et al.*, 1974; Adjepon-Yamoah *et al.*, 1974).

A thorough understanding of the metabolic pathway of a drug is important in characterizing its pharmacokinetic profile (Kwon, 2001). Figure 1.5 shows the biotransformation of drugs as an elimination pathway. Metabolism is usually catalysed by enzymes that can be found in most organs especially in the liver. If metabolism of a compound by one enzyme is blocked due to substrate saturation or by structural modifications, the compound can be metabolized by other types of enzymes (Kerns and Di, 2008). Drug metabolism or biotransformation is traditionally divided into two categories: Phase I and phase II reactions (Williams, 1959). Phase I metabolism results in the introduction of functional groups into molecules and hence it is also known as functionalization reaction. Phase II reactions are conjugation reactions with various endogenous compounds. Cytochrome P450 monooxygenase, and nitro and azo reductase are some of the main phase I enzymes, while important phase II enzymes include D-glucuronic acid, glutathione and sulfate transferase (Tsaioun and Kates, 2011). Phase I and II reactions normally produce more polar compounds with higher aqueous solubility.

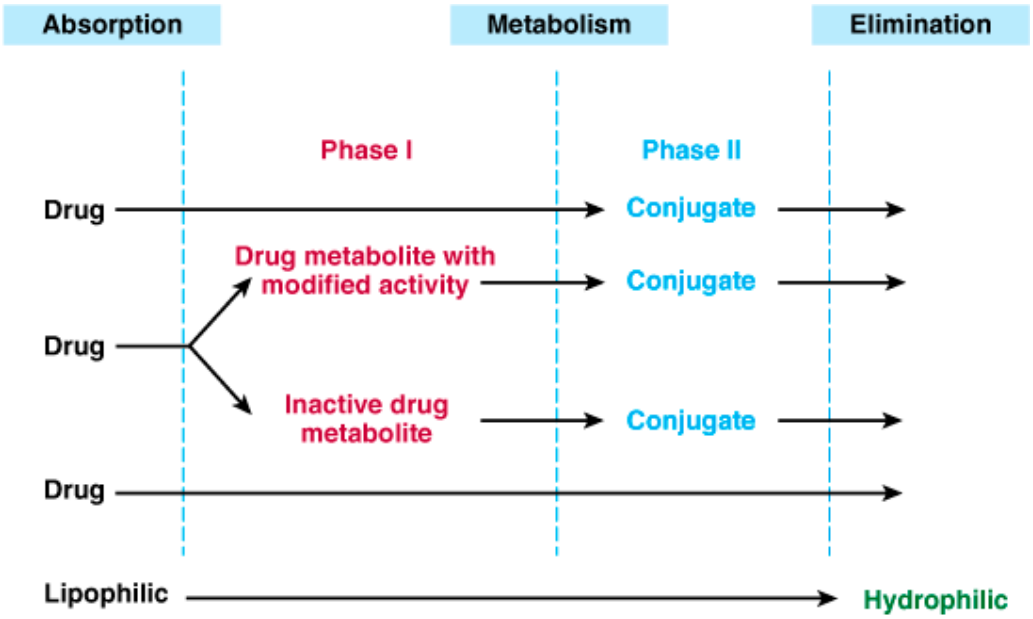

Figure. 1.5. Drug biotransformation (Katzung *et al.*, 2004).

1.5. Elimination by Membrane Transporters

Influx and efflux transporters are proteins expressed in cell membrane that have been shown to have a significant effect in the absorption, distribution and drug elimination. In the past ten years, there has been an enormous increase in the literature regarding the role of membrane transporters governing drug pharmacokinetics and response. An evaluation of the contribution of transporters to total tissue uptake and excretion is necessary to understand the drug disposition route (Giacomini *et al.*, 2010). Membrane transporters are classified according to their mode of transport, energy coupling mechanism, molecular phylogeny, and substrate specificity. Transporter categories include channels (e.g. Escherichia coli GlpF glycerol channel), primary active transporters (e.g. Lactococcus lactis LmrP multi drug efflux pump), ABC transporters (e.g. P-gp in humans and microorganisms), secondary transporters (e.g. E.coli LacY lactose permease) and group translocators (e.g. E.coli MtlA mannitol transporters) (Ren and Paulsen, 2005). Terada and co-workers have classified drug transporters into five main groups based mainly on their functions. There are: 1. Peptide transporters (PEPT),

2. Organic anion-transporting polypeptides (OATP), 3. Organic ion transporters (OAT, OCTN and OCT), 4. H^+/ organic cation antiporters (MATE) and 5. ATP-binding cassette (ABC) transporters (mainly P-gp, MRP1 and BCRP). The structures of these transporters, distribution in tissues and their roles are different. *In vivo* and *in vitro* techniques can be used to assess the character of transporters (Terada *et al.*, 2006).

Various transporters have been implicated in the clearance of several compounds and metabolites. Transporters are known to be partially responsible for drug concentration ratios in plasma and tissues, thus efficacy and toxicity. A big part of intact drug molecules and their metabolites are excreted into the bile by efflux transporters and passive diffusion into the bile channel (canaliculus) (Niemi *et al.*, 2011). Transporters can be found in all tissues but the four major locations that transporters operate significantly are intestinal epithelia, hepatocytes, kidney proximal tubules and blood-brain barrier (Giacomini *et al.*, 2010). Figure 1.6 illustrates a schematic representation of the important transporters and their positions in the membrane domain of different organs such as sinusoidal membranes of hepatocytes. As seen in this Figure, several uptake and efflux membrane transporters including apical ATP-dependant efflux pump (including P-gp, MRPs and BCRP), organic anion transporting polypeptide family (OATPs), ileal sodium-dependent bile acid transporter (ASBT), organic cation transporters (OCTs) family, peptide transporters (PEPTs), organic cation/carnitine transporters (OCTN), multidrug and toxin extrusion protein (MATE) and urate transporter govern the transport of compounds into and out of the cells.

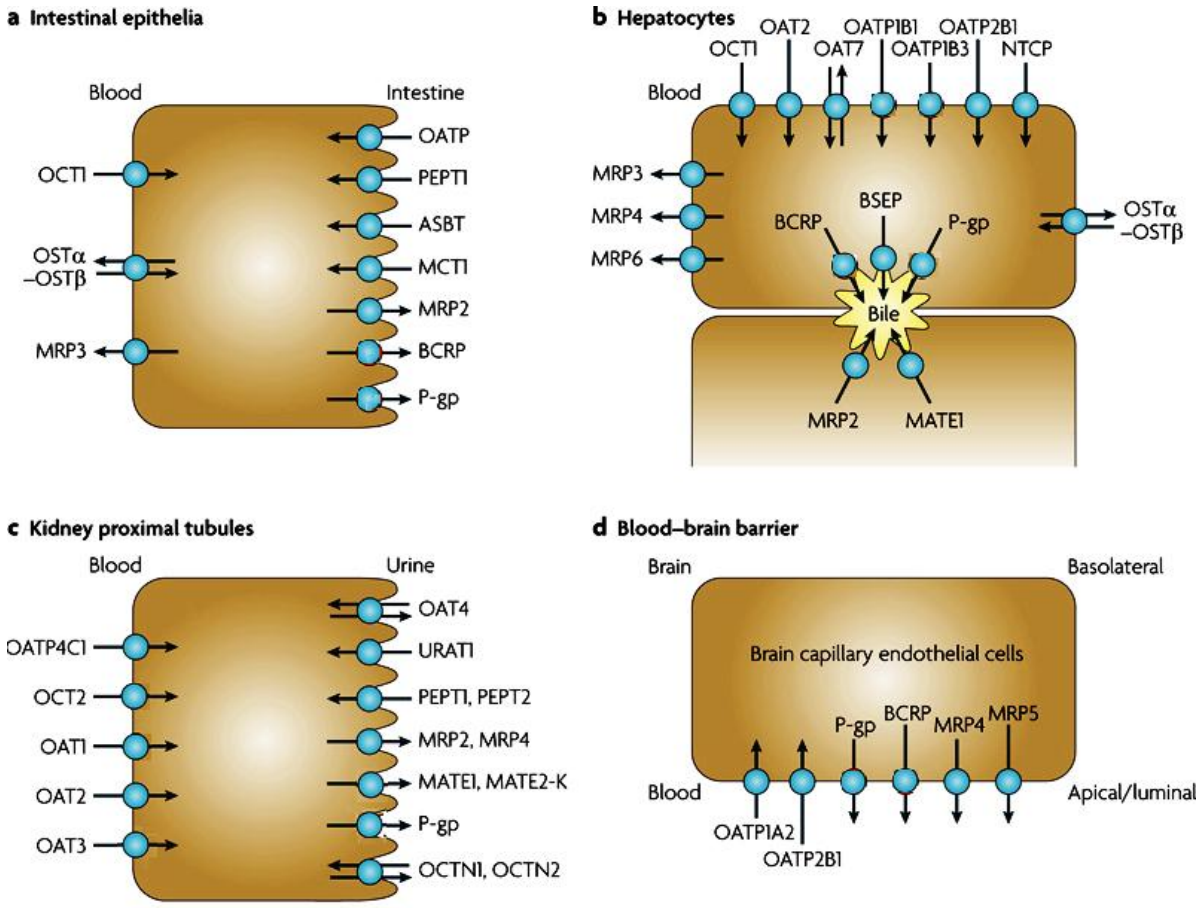

Figure. 1.6. The cartoon illustrates selected human transport proteins in plasma membrane domains of intestinal epithelia (a), hepatocytes (b), kidney proximal tubules (c) and brain capillary endothelial cells (d) (Adapted from Giacomini *et al.*, 2010)

1.5.1. Peptide Transporters (PEPT)

The currently known peptide transporters include peptide transporters 1 and 2 (Pept1 and Pept2) and peptide/histidine transporters 1 and 2 (PHT1 and PHT2). Studies showed that Pept1 is a low-affinity and high-capacity transport system for di and tripeptides (Leibach and Ganapathy, 1996). Conversely, Pept2 is a high affinity and low capacity transporter for di and tripeptides. The PHT1 and PHT2 transport di- and tri-peptides as well as histidine. These transporters are stereoselective as they show the affinity to L-enantiomers of amino acids (Doring *et al.*, 1998).

Pept1 was first cloned in rabbit intestinal epithelium membrane (Fei *et al.*, 1994). Pept2 for the first time was cloned from human kidney (Liu *et al.*, 1995). Pept1 and Pept2 can transport many peptides with different volumes and charges, but not long peptides with more than four peptide bonds (Daniel, 2004). These transporters are found mostly in the small intestine and kidney's proximal tubules and they mediate absorption of certain drugs e.g. cephalosporins and other beta-lactam antibiotics. There is no evidence of existence of these peptide transporters in blood brain barrier (BBB) (Han *et al.*, 1998). However, the expression of Pept1 was found with low levels in the liver, in addition to the major sites, small intestine and kidney (Liang *et al.*, 1995). Recently, the H^+-peptide cotransport has been established in the human bile duct epithelium cell line SK-ChA-1 (Knutter *et al.*, 2002).

Human PHT1 and PHT2 were found to be expressed at low levels in gastrointestinal tract and different tissues with mRNA expression throughout the gasterointestinal tract. In addition, the mRNA expression was also demonstrated in the liver, brain, colon, heart, kidney, lung, ovary, pancreas, placenta, prostate, spleen and testis (Herrera-Ruiz *et al.*, 2001).

In the past decade, amino acid modifications have been used in the design of prodrugs to allow for PEPT1 and PEPT2 intestinal absorption of weakly absorbed drugs such as antiviral agents levovirin and azidothimidine, and anticancer drugs gemcitabine and floxuridine (Sugawara *et al.*, 2000; Li *et al.*, 2006).

1.5.2. Organic Anion-Transporting Polypeptides (OATP)

OATP is a family of membrane transporters that mediate the cellular uptake of endogenous substrates and drugs. The importance of OATPs in excretion has been shown by different studies (Cvetkovic *et al.*, 1999; Mikkaichi *et al.*, 2004; Kim, 2003). The human OATP family consists of 11 members: OATP1A2, 1B1, 1B3, 1C1, 2A1, 2B1, 3A1, 4A1, 4C1, 5A1 and 6A1 (Hagenbuch and Meier, 2003). As seen in Figure 1.6, members of this family can be found in sinusoidal (basolateral) membrane of hepatocytes, basolateral membrane of proximal tubules, and apical (luminal) side of the blood-brain barrier and intestinal epithelia. Certain OATP isoforms are selectively involved in hepatic uptake of hydrophobic anions from the

plasma (Taft, 2009). Although the role of OATPs in renal (Sekine *et al.*, 2006) and hepatic excretion (Nozawa *et al.*, 2005) as well as uptake across the blood-brain barrier (Gao *et al.*, 2000) and gastrointestinal tract (Sai *et al.*, 2006) has been demonstrated, their importance in pharmacokinetics is still not fully understood (Glaeser and Kim, 2006).

Despite the title, OATP substrates are not limited to organic anions, but also include cations as well as neutral and zwitterionic compounds (Niemi *et al.*, 2011). The OATP family members mediate the sodium ion co-transport of various organic agents including organic dyes, bile salts, steroid conjugates and thyroid hormones. In rat, the organic anion transporting polypeptides Oatp1, Oatp2 and Oatp4 have been indicated as the main sodium independent uptake proteins (Kullak-Ublick *et al.*, 2000).

OATP structure is a protein with twelve transmembrane domains (Hagenbuch and Gui, 2008). The first of the organic anion-transporting polypeptides OATP1A2 (OATP1) was originally cloned from a human kidney cDNA library (Lu *et al.*, 1996). Later, OATP1A2 was cloned from rat liver and since then, several different forms of OATPs in human and rodents have been discovered (Jacquemin *et al.*, 1994). For instance, OATP1B1 was cloned independently by different laboratories (Tirona *et al.*, 2001; Hsiang et, 1999; Konig *et al.*, 2000a; Abe *et al.*, 1999). OATP1B3 was also cloned from human liver (Abe *et al.*, 2001; Konig *et al.*, 2000b). OATP1B3 is mainly expressed in the basolateral membrane of hepatocytes (Abe *et al.*, 2001), but it has also been detected in certain cancer cell lines and tissues (Abe *et al.*, 2001). Over the last two decades the impact on drug pharmacokinetics of the organic anion transporting polypeptides (OATPs: OATP-1B1, 1B3 and 2B1), expressed on the sinusoidal membrane of the hepatocyte, has been increasingly recognized.

Human OATP1B1 (also known as OATP2) is a liver specific transporter that is expressed on the sinusoidal membrane of human hepatocytes and mediates the hepatic uptake of many endogenous compounds. The substrate specificity of OATP1B1 is closely comparable to OATP1A2 and both can transport drugs such as eicosanoids, benzylpenicillin, methotrexate, rifampin, pravastatin, rosuvastatin and cerivastatin (Glaeser and Kim, 2006). Apart from hepatocytes, OATP1A2 is

expressed in various tissues including brain and kidneys. Moreover, OATP1A2 can facilitate the entry of its substrates through the duodenal wall into circulation (Glaeser *et al.*, 2007). Regarding the acidic, basic and neural character of substrates, OATP1A2 possesses perhaps the broadest spectrum among the members of the superfamily (You and Morris 2007).

OATP1B3 has a significant substrate overlap with OATP1B1 (Karlgren *et al.*, 2012a). However, OATP1B3 is also able to transport oligopeptide hormones such as cholecystokinin 8 (Ismair *et al.*, 2001) and digoxin (Kullak-Ublick *et al.*, 2001), although the latter has been disputed (Taub *et al.*, 2011). Unlike OATP1B1, OATP1B3 has been implied in the transport of angiotensin II receptor antagonist, telmisartan, and its glucuronide conjugate (Abe *et al.*, 1999) as well as mediating the cellular uptake of opioid peptide II, digoxin and ouabain (Kullak-Ublick *et al.*, 2001). The importance of OATP1B1 and OATP1B3 in hepatic transport has been explained by recent studies by Fenner and co-workers indicating that OATP1B-mediated transport can be the rate-determining step of hepatobiliary drug clearance (Fenner *et al.*, 2012).

In addition to drug clearance role, recent studies have suggested that overexpression of OATP1A2, OATP1B1 and OATP1B3 in pancreatic cancer tissues (Kounnis *et al.*, 2011) as well as in ovarian cancer cells (Svoboda *et al.*, 2011) may be exploited in the design of novel targeted cancer therapy (Sainis *et al.*, 2010). This is particularly important in light of the increasing global burden of cancer. GLOBALCAN 2008 (Ferlay *et al.*, 2010) reported over 12.7 million cancer cases and 7.6 million cancer deaths are estimated to have occurred in 2008 and deaths from cancer worldwide are projected to continue rising with an estimated 13.1 million deaths in 2030 (Jemal *et al.*, 2011).

1.5.3. Organic Ion Transporters (OAT, OCTN and OCT)

Organic anion and cation transporters (OATs and OCTs) and organic cation/carnitine transporter (OCTN) superfamily are members of the solute carrier family, subfamily 22 (SLC22). These transmembrane proteins are largely expressed in excretory organs such as kidney and liver, as a major component of the human

xenobiotic excretion machinery. In the liver, these uptake transporters play important role in the initial sinusoidal influx of drugs into hepatocytes (van Montfoort *et al.*, 2003) (see Figure 1.6). These transporters have wide substrate specificities for a range of exogenous and endogenous substrates including many commonly used drugs, antibiotics, anti-hypertensives, and anti-inflammatories, among others (Leabman *et al.*, 2003).

In kidneys, organic cation transporters mediate the transport of small organic cation such as tetraethylammonium. OCT1 was the first discovered OCT from rat kidneys in 1994 (Grundemann *et al.*, 1994). In humans, OCT1 is expressed at extremely low levels in the kidney and is mainly found in the liver (Motohashi *et al.*, 2002). As seen in Figures 1.4 and 1.6, OCT1 can be found abundantly in hepatocytes and may be seen as the most important transporter for distribution of cationic compounds into the liver from sinusoidal membrane (Nies *et al.*, 2009). OCT2 was isolated from the rat kidney using cDNA cloning of the OCT1 sequence (Okuda *et al.*, 1996). OCT2 is generally considered to be a kidney transporter, though mRNA is expressed at low levels in other tissues such as spleen, placenta, small intestine and brain (Gorboulev *et al.*, 1997). OCT3 has the widest tissue distribution of the OCTs and its protein expression has been confirmed on the basolateral membrane of hepatocytes (Nies *et al.*, 2009), the basal membranes of trophoblasts (Sata *et al.*, 2005), the apical membrane of enterocytes (Muller *et al.*, 2005) and the luminal membrane of lung epithelial cells (Lips *et al.*, 2005). Substrates for OCT1-3 include a wide range of structurally unrelated organic cations, including many drugs. An extensive list of OCT1-3 substrates and inhibitors has been provided in a recent review on the importance of organic cation transporters in drug therapy (Nies *et al.*, 2011). Among these substrates are catecholamines, monoamine neurotransmitters and several antiviral drugs.

OATs are fairy well-studied organic anion transporters and are mainly expressed in excretory organs, especially kidney for the uptake of organic anions from the blood to renal tubule cells (see Figure 1.6). OATs are membrane proteins with 12 putative membrane-spanning domains and function as sodium-independent exchangers or facilitators. OATs mediate the influx of a wide range of organic anions including inorganic ions (e.g. CL^- and $HCO3^-$), endogenous (e.g. cyclic nucleotides,

prostaglandins, urate, dicarboxylates) and exogenous anions (various anionic drugs and environmental substances) (Sekine *et al.*, 2000). In comparison with OATPs, substrates of OAT have been suggested to be generally lower molecular weight (Roth *et al.*, 2012). The transport mechanism of OAT1 and OAT3 is known to be indirectly sodium-dependent and involves a 'tertiary active transport' mechanism to move organic anions across the basolateral membrane into the proximal tubule cells. The primary active Na^+ and K^+-ATPase located on the basolateral membrane pumps Na^+ from intracellular to extracellular space to maintain a Na^+ gradient (Glaeser and Kim, 2006). This is used by the secondary active Na^+-dicarboxylate cotransporter to maintain a high intracellular concentration of α-ketoglutarate, which is used to drive uptake of other organic anions by OAT1 and OAT3. Several studies have revealed that rat Oat1 transports a broad spectrum of substrates (Glaeser and Kim, 2006). Endogenous organic anions such as prostaglandins, cyclic nucleotides, folates (Sekine *et al.*, 1997) and some xenobiotics such as beta-lactam antibiotics (Jariyawat *et al.*, 1999; Leabman *et al.*, 2003), NSAIDs (Apiwattanakul *et al.*, 1999) as well as many antiviral drugs (Cihlar *et al.*, 1999; Wada *et al.*, 2000) are examples of compounds transports by rat Oat1. Human OAT1 also transports adefovir, cidofovir, zidovudine (AZT), acylclovir and ganciclovir (Cihlar *et al.*, 1999; Ho *et al.*, 2000).

OAT2 mRNA has the highest expression levels in the liver with lower levels also seen in kidney (Sekine *et al.*, 1998; Sun *et al.*, 2001; Hilgendorf *et al.*, 2007). Human OAT3 is exclusively expressed in the basolateral membrane of the proximal tubule cells of kidneys (Cha *et al.*, 2001; Sun *et al.*, 2001) while in rat, Oat3 is most abundantly expressed in liver and to lesser extent in kidney and brain (Kusuhara *et al.*, 1999). OAT4 mRNA is expressed in kidney and placenta (Bleasby *et al.*, 2006). OAT5 expression in human is not well studied, although Northern blot analysis demonstrates mRNA expression in the liver (Sun *et al.*, 2001). OAT7 has been shown to be exclusively expressed in the liver, where its expression has been localized to the basolateral membrane of hepatocytes (Shin *et al.*, 2007). OAT10 mRNA has the highest expression levels in the kidney followed by brain, heart, small intestine and colon (Bahn *et al.*, 2008). URAT1 is expressed in kidney and it is the only member of the OAT family for which mutations have been linked to a disease (Enomoto *et al.*, 2002).

Carnitine is an essential zwitterion cofactor that plays an important role in the metabolism of lipids and subsequently in the production of energy. Carnitine absorption is via small intestine with the help of Organic Cation/Carnitine transporter 2 (OCTN2), which is located on the brush border membrane (Elimrani *et al.*, 2003). OCTN2 transports organic cations without involving Na^+, but it transports carnitine only in the presence of Na^+. Wu and colleges found that rat OCTN1 is expressed in a wide variety of rat tissues and organs such as intestine, liver, kidney, heart and brain (Wu *et al.*, 2000). OCTN2 is also expressed in the heart, kidney, placenta and brain (Wu *et al.*, 1999). There is no evidence of presence of OCTN2 in human liver while it is strongly expressed in rat liver (Tamai *et al.*, 1998).

1.5.4. H^+/ Organic Cation Antiporter (MATE)

Multidrug and toxin extrusion transporters (MATE) mediate cellular efflux of a variety of organic cations, including many drugs (Lickteig *et al.*, 2008). MATE1, which functions as drug/sodium antiporter, is the first example of Na^+-coupled multidrug efflux transporter (Morita *et al.*, 2000). The MATE are protein transporters which are primarily expressed in the kidney and liver, localized at the apical membranes of the renal tubules and bile canaliculi (Motohashi and Inui, 2013; Motohashi *et al.*, 2013). MATE1 has been isolated as an H^+/organic cation antiporter located at the renal brush-border membranes (Asaka *et al.*, 2007). MATE1 can transport zwitterionic drugs such as fexofenadine and levofloxacin, as well as organic cation drugs such as metformin and cimetidine (Terada *et al.*, 2006; Masuda *et al.*, 2006).

In rat, apart from kidney, MATE1 also expressed abundantly in the placenta, slightly in the spleen, but not expressed in the liver (Terada *et al.*, 2006). Rat multidrug and toxin extrusion (MATE1) transporter is expressed in kidney, but not in the liver (Ohta *et al.*, 2006; Masuda *et al.*, 2006). In humans, MATE1 mRNA levels are highest in the liver, and are localized to the canalicular membrane of hepatocytes. MATE1 mRNA expression is also high in the kidneys, where it is localized to the apical membrane of the renal tubule. Similarly, MATE2 mRNA

levels are by far at their highest in the kidneys, while relatively low in most other tissues (Lickteig *et al.*, 2008).

1.5.5. ABC Transporters

ATP-binding cassette (ABC) transporters are transmembrane proteins that utilize the energy of adenosine triphosphate (ATP) binding and hydrolysis to carry out certain biological processes including translocation of various substrates across membranes. These are mainly efflux transporters that help export compounds out of the cells (Massey *et al.*, 2014). Amongst the largest transporter superfamilies, these transporters may be found in all known organisms and around 1100 various transporters belong to this group (You and Morris, 2007). Figure 1.6 illustrates several members of ABC transporters in brain, kidney, intestine and liver. In the liver, the ABC transporters MRP2, BCRP, P-gp and BSEP (ABCB11 and also known as sPgp (sister of P-glycoprotein)) are found in the canalicular membrane of hepatocytes exporting the substrates into the bile. Other members of ABC transporter family, including MRP3, MRP4 and MRP6, are distributed in sinusoidal membrane and they export the substrates from hepatocytes back into the blood. ABC transporters can be found in many normal tissues with an important role in drug elimination or other biological processes.

Genetic defects in some of the ABC transporters may result in a disease; mutations in up to 14 mammalian ABC transporters (out of 48 ABC genes) have been associated with disease states (Borst and Elferink, 2002). For example, dysfunction of ABCB2 transporter results in immune deficiency problems and dysfunction of ABCC2 results in Dublin-Johnson syndrome (Gottesman and Ambudkar, 2001).

These transporters are further categorised into seven distinct subfamilies of proteins using phylogenetic analysis. The subfamilies include: ABCA (12 members), ABCB (11 members), ABCC (12 members) ABCD (4 members), ABCE, ABCF (3 members) and ABCG (1 member) (Hennessy and Spiers, 2007). The best-studied proteins of this family include P-gp (ABCB1) also known as MDR1 due to its ability to produce multiple drug resistance in cancer cells, and the sulphonylurea receptor (SUR) subfamily encoded by members of ABCC genes that is involved in

regulating insulin secretion in β-cells of the pancreas (Dassa and Bouige, 2001). Others include the ABCC subfamily which encodes the cystic fibrosis transmembrane conductance regulator (CFTR) protein that plays a part in exocrine secretions of chloride (Dean, 2002; Dassa and Bouige, 2001). A number of these proteins including MRP1, BCRP and P-gp are reported to be overexpressed in malignant cells thus causing these cells to be resistant to drug therapy, hence the multidrug resistance (MDR) terminology.

In eukaryotic cells, ABC transporters usually direct molecules from the cytoplasm to the outside of the cell (Dean, 2002) with the main function of transporting xenobiotic compounds out of the cell for transport to other areas of the body or for excretion. On the other hand, ABC transporters in prokaryotic cells can be either an importer or exporter of compounds. Bacterial importers are important for the cell survival and typically important substrates such as iron, inorganic ions as well as peptides and amino acids. Substances requiring removal from prokaryotic cells include cell wall components such as liposaccharides and toxins involved in pathogens e.g. haemolysin (Davidson *et al.*, 2008).

Structurally, ABC transporters consist of two distinct domains, the nucleotide binding domain (NBD) and the transmembrane domain (TMD). A typical ABC transporter may have two TMD domains and two NBD domains (Higgins, 2001). The TMD of various ABC transporters is diverse and could contain 6–11 membrane-spanning α-helices and provides the specificity for the substrate in order to function as the route for molecules to cross the membrane. The NBDs of the protein, also known as the ATP-binding domain, can be found in the cytoplasm and are consequently hydrophilic in nature (Dean, 2002). These domains help transfer the energy needed to transport the substrate across the membrane (Dean, 2002; Ambudkar *et al.*, 2003). NBD consists of two subdomains: 1. ‘the catalytic core domain’ that includes walker motif A and walker motif B with a dodecapeptide part that connects the two walker motifs, and 2. a smaller, structurally diverse α-helical subdomain that contains the ABC signature motif. ABC transporter proteins bind ATP through their NBDs and use the energy derived from this to transfer molecules across cell membranes. A glutamine residue residing in a flexible loop called Q loop that connects the TMD and NBD is presumed to be involved in the interaction

of the NBD and TMD, particularly in the coupling of nucleotide hydrolysis to the conformational changes of the TMD during substrate translocation. The H motif or switch region contains a highly conserved histidine residue that is also important in the interaction of the NBD domain with ATP.

1.5.5.1. ABC Transporters in Multidrug Resistance

During cancer treatment, tumour cells can become resistant to chemotherapy due to increased excretion of drugs out of tumour cells or target proteins (Dean, 2002). Pathways such as these can lead to multidrug resistance (MDR) thus contributing to the failure of chemotherapy in malignant diseases. Multidrug resistance is the term given to describe tumours developing resistance to two or more chemotherapeutic drugs. This is the net result of the overexpression of membrane transporters that actively remove toxic chemotherapeutic agents out of tumour cells (Sarkadi *et al.*, 2006). ABC transporters have been widely associated with resistance and the ABC genes ABCB1 (encoding P-gp), ABCC1 (encoding MRP1) and ABCG2 (encoding BCRP) are the main genes that can be upregulated in cancerous cells. MRP1 is expressed in epithelial cells and in non-malignant cells it plays a role in protecting kidney tissues, bone marrow and the intestinal mucosa from xenobiotics as well as contributing to the removal of drugs from the cerebrospinal fluid (Schinkel and Jonker, 2003). Moreover, MRP1 confers drug resistance to a range of cancer drugs and transports conjugates of hydrophobic drugs as well as organic anions (Schinkel and Jonker, 2003). P-glycoprotein (P-gp) was one of the first ABC transporters to be associated with resistance (Leslie *et al.*, 2009) and led to the discovery of other genes in the ABC transporter family involved in multidrug resistance. P-gp is highly expressed in cancerous tissues and it is reported to be involved in cancers of the liver, colon and kidney tissues (Schinkel and Jonker, 2003). Breast Cancer Resistance Protein (BCRP) was discovered after analysis of mitoxantrone-resistant cell lines that did not over-express P-gp or MRP1 by Doyle et al (1998). It was first cloned from a multidrug-resistant breast cancer cell line, hence the name.

In addition to chemotherapeutic agents P-gp, BCRP and MRP1 also actively transport non-cytotoxic drugs and xenobiotics (Matsson *et al.*, 2009; Sharom, 2008;

Mao and Unadkat, 2005), thereby affecting the pharmacokinetics and tissue distribution of these drugs. Table 1.1 gives a summery description of these transporters.

Table 1.1. Properties of ABC transporters

Comm-on names	Systematic name	Tissue localisation	Substrates	Inhibitors
P-gp/ MDR1	ABCB1	Apical membranes of the intestine, liver, kidney, placenta and blood brain barrier (BBB)	***Cancer drugs***: Anthracyclines, vinca alkaloids, taxanes, captothesins, anthracenes and epipodophyllotoxins ***Non-cancer drugs***: Digoxin	First generation inhibitors like verapamil; Second generation inhibitors such as valspodar and third generation inhibitors like Elacridar (GF120918)
MRP1	ABCC1	Basolateral membranes of all tissues, and possibly apical membrane of the BBB	***Cancer drugs***: anthracyclines, vinca alkaloids, captothesins, epipodophyllotoxins and methotrexate ***Other compounds***: Glutathione, sulphate and glucuronide conjugates	BSO, flavonoids, HIV protease inhibitors, non- HIV protease inhibitors, PAK-104P and MK571
BCRP	ABCG2	Apical membranes in the intestines, liver, immature stem cells, the brain, mammary glands and placenta	***Cancer drugs***: anthracyclines, captothesins, epipodophyllotoxins, mitoxantrone, flavopiridol, methotrexate and bisantrene ***Other compounds***:, drug and metabolite conjugates, food carcinogens like PhiP and other drugs	Flavonoids, fungal toxins like FTC, calcium channel blockers and tyrosine kinase inhibitors

Data from Sharom, 2008; van Herwaarden and Schinkel, 2006 and Gottesman *et al.,* 2002

Apart from their role in MDR, transporter proteins encoded by ABCB1, ABCC family (mainly MRP2) and ABCG2 have major functions in the pharmacokinetics and tissue distribution of different drugs. As can be seen in Figure 1.6, P-gp, BCRP and MRP2 are located in the apical membrane of intestinal epithelia and export the substrate compounds from epithelial cells back into the lumen, while MRP3 is located in the basolateral membrane and transports its substrates from cytoplasm into the blood. The main ABC transporters in the kidney are P-gp, MRP2 and MRP4, with an efflux role for active secretion of their substrates. P-gp, BCRP and MRP2 are also involved in bile secretion through efflux of their substrates in the canalicular membrane. P-gp, BCRP, MRP4 and MRP5 are the main ABC transporters responsible for the efflux of compounds from the brain. Below is a description of these ABC transporters in terms of their structure, binding and efflux mechanisms, substrates, inhibitors and polymorphisms.

1.5.5.1.1. P-glycoprotein (ABCB1 Subfamily, MDR)

The schematic diagram of P-gp can be seen in Figure 1.7. This protein consists of 1280 amino acids forming 12 transmembrane segments. P-gp has an exceptionally wide range of substrate specificity for cationic and lipophilic drugs. Apart from drugs, P-gp as a strong efflux pump is able to export a number of structurally diverse compounds including anthracyclines, epipodophyllotoxins and vinca alkyloids (Eckford and Sharom, 2009).

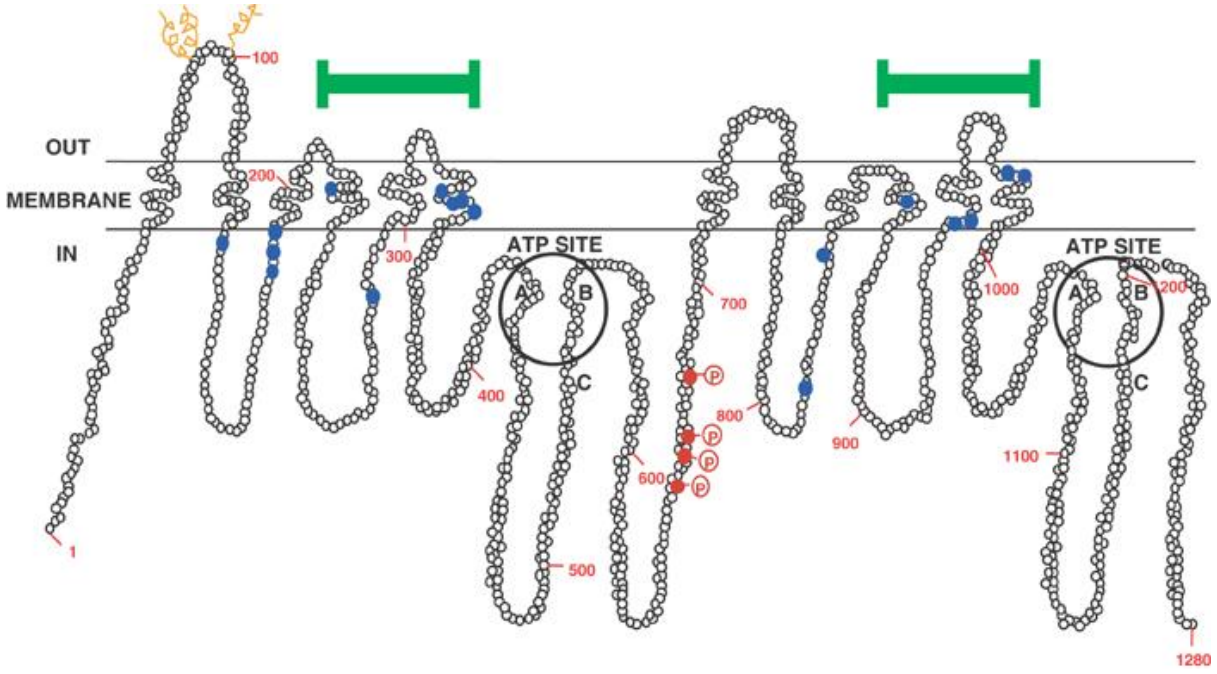

Figure 1.7. Schematic diagram showing the structure of human P-gp with 1280 amino acids and 12 transmembrane segments. Each loop in this topological view represents an amino acid residue (Adapted from Gottesman and Pastan, 1988).

P-gp is expressed at many physiological barriers such as the intestinal epithelium, hepatocytes, renal proximal tubular cells, pancreatic and bile ductules, adrenal gland and the endothelial capillaries of the brain comprising the blood brain barrier (Kim *et al.*, 1998; Thiebaut *et al.*, 1987; Croop *et al.*, 1989). This transport protein plays a significant role in different steps of absorption, distribution, metabolism and elimination of many compounds including anticancer drugs (Schinkel *et al.*, 1995; Leveque and Jehl, 1995; Relling, 1996). In the membrane of hepatocytes, where P-gp is mostly expressed, P-gp is involved in the efflux of xenobiotics into the bile (Yu *et al.*, 2010). In the gastrointestinal tract, P-gp pumps out the substrates into the gastric lumen; in such a case, the agents cannot access the portal vein to reach the systemic circulation (Schinkel *et al.*, 1997). Therefore, P-gp can reduce the absorption and oral bioavailability of the substrate drugs. Moreover, it can be found in testis barrier (Melaine *et al.*, 2002), blood brain barrier cells (Beaulieu *et al.*, 1997), blood mammary tissue barrier (Edwards *et al.*, 2005), blood-inner ear barrier (Saito *et al.*, 1997), placenta (Gil *et al.*, 2005) and endometrium of pregnant women (Arceci *et al.*, 1988). A natural function of P-gp is that it prevents harmful chemicals or foreign compounds (xenobiotics) including drug molecules from getting into the brain and the placenta (Lin and Yamazaki, 2003).

P-gp is highly overexpressed in tumour cells and is able to bind and transport many chemically and structurally unrelated drug molecules thus explaining its MDR ability in cancer chemotherapy (Gottesman and Ambudkar, 2001). As a consequence of P-gp blockage, e.g. in the presence of inhibitors, the intracellular accumulation of the substrate drugs (chemotherapeutic agents) will increase which may result in excessive toxicity of these drugs. However, the reduction of chemotherapeutic dose is not a solution as it will reduce the overall efficacy (Wacher *et al.*, 1995; McDevitt and Callaaghan, 2007; Wandel *et al.*, 1999). An example of this situation is when a drug molecule such as digoxin, which is a P-gp substrate, is accumulated in the liver and kidney as a result of P-gp inhibitors preventing the biliary and renal elimination of digoxin by active secretion with the aid of P-gp efflux system (Hennessy and Spiers, 2007).

P-gp has a promiscuous binding site that can accept a wide range of substrates of varying chemically unrelated chemical structures. The weight range of P-gp substrates can be very broad and vary from a MW of 250 to 1850 Da. Besides, the substrate molecules can be acidic, zwitterionic, uncharged or positively charged (Schinkel *et al.*, 1997). Moreover, substrates can be amphipathic or hydrophobic (Kerns and Di, 2008). In terms of the modulators of this multispecific transporter, not only pharmaceutical drugs but also herbal products and some food components can affect the function of P-gp as a transporter. It is therefore advisable that in drug discovery, when a drug candidate is found to be a P-gp substrate, structure modifications are applied to reduce the P-gp activity, leading to a better therapeutic effect with less complications such as drug-drug interactions in drug discovery projects (Kerns and Di, 2008).

The structure of human P-gp was first elucidated by electron microscopy (Rosenberg *et al.*, 1997) and image analysis. P-gp was reported as having a central core with an opening to the extracellular side of the membrane but is closed towards the cytoplasm. Recently, Aller et al reported a medium resolution (3.8-4.4 Å) X-ray structure of P-gp that supported previous claims about the structure of P-gp and revealed tentative binding sites for drug compounds (Aller *et al.*, 2009). The study proposed a detailedstructure for mouse P-gp which has 87% sequence identity to human P-gp. In addition to the structure of apo P-gp at 3.8 angstroms,

two structures of P-gp co-crystallised with cyclic peptide inhibitors cyclic-tris-(R)-valineselenazole (QZ59-RRR) and cyclic-tris-(S)-valineselenazole (QZ59-SSS) were also determined. The structures showed distinct drug-binding sites in the internal cavity capable of stereoselectivity that is based on hydrophobic and aromatic interactions. The structure of apo P-gp reveals a large internal cavity that is approximately 6000 angstroms cubed, and a big gap of 30 angstrom between the two nucleotide-binding domains (Figure 1.8). In agreement with previous theories (Rosenberg *et al.*, 1997, Higgins and Gottesman, 1992), the apo and drug-bound P-gp structures in Aller's work indicate portals open to the cytoplasm and the inner leaflet of the lipid bilayer for drug entry. The inward-facing conformation represents an initial stage of the transport cycle that is competent for drug binding (Figure 1.8).

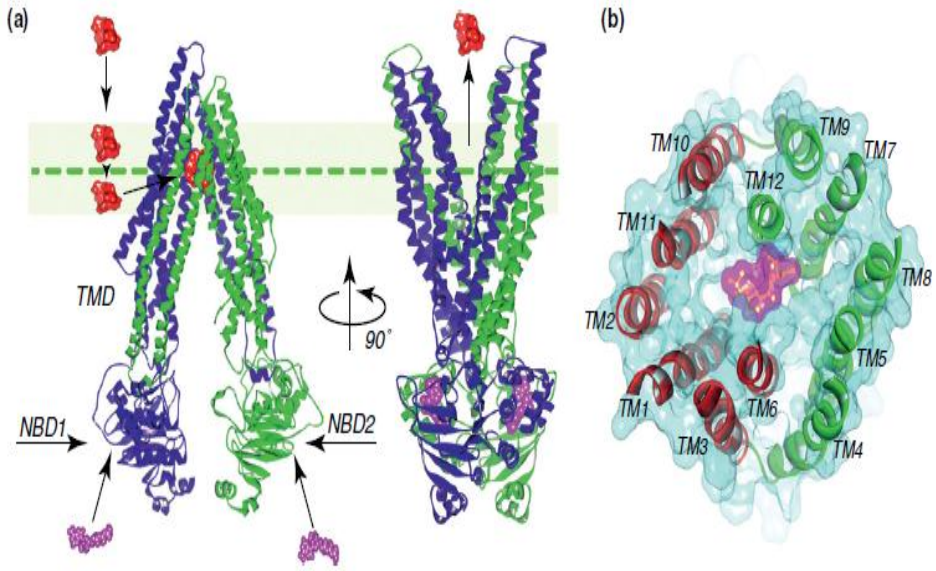

Figure 1.8. (a) P-gp structure and efflux activity; substrates are in red while ATP is in magenta. (b) Ligand-binding site on the transmembrane domain of P-gp (adapted from Chen *et al.*, 2012).

The X-ray crystal structures proposed by Aller gave some useful information regarding the amino acid residues involved in substrate binding to P-gp. The crystal structure PDB Code (3G60) showed one molecule of QZ59-RRR bound to the middle site in the binding pocket and two molecules of QZ59-SSS bound at upper and lower sites which are overlapping the middle site. This showed that P-gp can

bind to two drug molecules at the same time and confirmed the diverse and polyspecific nature of P-gp (Aller *et al.*, 2009).

According to Aller and co-workers, the binding pocket of P-gp includes the transmembrane helices 1, 6, 7 and 12 which mainly consist of hydrophobic and aromatic residues. These included phenylalanine (Phe) and tyrosine (Tyr) residues in addition to the aromatic and aliphatic residues serine, Threonine and Glutamine (Ser, Thr, Gln). Despite these key attributes being made available, questions have been raised about the absence of ATP in the structure and the fact that the structures do not appear to undergo conformational changes upon drug binding (Gottesman *et al.*, 2009).

Substrates of P-gp mainly interact with the protein by hydrophobic interactions, π-π stacking and van der Waals forces. The P-gp X-ray crystal structure also shows this as the cyclic peptide inhibitors bind to P-gp through hydrophobic aromatic side residues (Aller *et al.*, 2009). Studies have also demonstrated that P-gp is a flexible molecule that can alter its conformation in order for substrate entry. These findings led to a proposed induced-fit mechanism for drug binding to P-gp, in which the substrate enters the large binding pocket and both drug and protein modify their shape to generate more favourable contacts unique to that substrate (Alonso *et al.*, 2006). This mechanism is supported by the X-ray structure of P-gp, where each of the ligands bound to P-gp interact with the protein at different or the same overlapping amino acid residues. Recent site-directed mutagenesis studies have provided evidence that each substrate can bind to more than one site and all sites are capable of transport function (Chufan *et al.*, 2013).

1.5.5.1.2. Multidrug Resistance-Associated Protein (MRP, ABCC Subfamily)

Multidrug resistance-associated protein consists of ABCC1, ABCC2, ABCC3, ABCC4, ABCC5, ABCC6, ABCC10, ABCC11 and ABCC12 (You and Morris, 2007). All of these MRPs act as efflux pump.

Many compounds including glutathione conjugates were identified as MRPs substrates including LTD4, S-glutathionyl 2,4-dinitrobenzene (DNP-SG), 17β-

glucuronosyl estradiol, lithocholyltaurine 3-sulfate, oxidized glutathione and bilirubin glucuronosides (Jedlitschky *et al.*, 1996). Furthermore, numerous unconjugated amphiphilic anions are transported by ABCC1. Examples are folate and its antimetabolite methotrexate (Hooijberg *et al.*, 1999). Its function as a pump for cytostatic agents, confers resistance a broad range of anti-cancer drugs. MRP1 is mostly found in the lung, testis, kidney, and macrophages. MRP1 shares a similar distribution pattern with MRP2, which holds the role of excretion and detoxification of endogenous and xenobiotic anions in the bile (Nies *et al.*, 2007). However, localization of MRP1 makes its role more to protect the cells from toxic effects of endogenous and xenobiotic anions rather than excretion (Bakos and Homolya, 2007).

ABCC2 (MRP2) is an efflux transporter which transports sulphate conjugates, glucuronide and glutathione of many compounds and xenobiotics (Jansen *et al.*, 1985). This transporter abundantly exists in canalicular membrane of liver and plays crucial role in the biliary transport of anionic conjugates. Studies in mutant rats indicated that the lack of functional MRP2 leads to deficiency in the secretion of anionic conjugates into bile (Hosokawa *et al.*, 1992). MRP2 has a crucial role in the biliary secretion of many endogenous and exogenous compounds (Morikawa, *et al.*, 2000) and down-regulation of MRP2 expression leads to impaired biliary excretion of amphiphilic anionic conjugates in the rat models of cholestasis (Trauner *et al.*, 1997).

ABCC3 (MRP3) can transport a wide range of endogenous and exogenous substrates (mainly conjugated organic anions) to blood circulation. As shown in Figures 1.4 and 1.6, unlike MRP2, this transporter is mostly expressed at the basolateral membranes of liver and intestine (Ehrhardt and Kim, 2008). Studies in mutant rats with chronic conjugated hyperbilirubinemia, which are unable to secrete bilirubin glucuronosides into bile shows that hepatic MRP3 expression is inducible but appears to be constitutive in other organ (Hirohashi *et al.*, 1998; Fernández-Barrena *et al.*, 2012). MRP3 may function as a “backup” transporter for amphipathic conjugates in cholestatic conditions. It may have a role in detoxification of hepatocytes by extruding bile acids and other conjugates into sinusoidal blood.

ABCC4 (MRP4) is characterized as an ATP-dependent organic anion transporter. Nucleoside monophosphate analogues were the first substrates that were discovered for MRP4 (Schuetz *et al.*, 1999). In addition, transport of the prostaglandins PGE1 and PGE2 is mediated by MRP4 (Reid *et al.*, 2003). MRP4 is acquired in basolateral as well as in apical membrane localizations. MRP4 was found in apical membrane of proximal tubule epithelial of human cells (van Aubel *et al.*, 2002) and rat kidney (Denk *et al.*, 2004). MRP4 was demonstrated in the basolateral membrane in human, rat and mouse hepatocytes (Denk *et al.*, 2004 and Rius *et al.*, 2003) (See Figures 1.4 and 1.6).

ABCC5 (MRP5), similar to MRP4 may be found either in basolateral or apical membrane. In intact human cells, MRP5 was able to mediate efflux of the anionic dye fluorescein diacetate with ATP consumption (McAleer *et al.*, 1999). The ABCC6 (MRP6) protein is detectable in liver and kidney, in the basolateral membrane of rat (Madon *et al.*, 2000) and in hepatocytes in human (Keppler *et al.*, 2001). ABCC10, ABCC11, and ABCC12 are recently identified members of the MRP family that are at relatively early stages of investigation. ABCC10 and ABCC11 are lipophilic anion pumps that are able to confer resistance to chemotherapeutic agents. ABCC11 is an efflux pump that is able to transport cyclic nucleotides (Guo *et al.*, 2003). It is also able to transport leukotriene C4 (LTC4), 2,4-dinitrophenyl glutathione (DNP-SG), estradiol 17-β-D-glucuronide (E217βG), monoanionic bile salts cholyglycine and cholyltaurine, folate and antimetabolite methotrexate, steroid sulphates E13S and DHEAS (Chen *et al.*, 2005). In human, ABCC11 is localized in the cerebral cortex of neurons. A recent study on localization of ABCC proteins has shown the expression of ABCC11 in Sertoli (rat testis cells) (Klein *et al.*, 2014). The human genes and transmembrane helices of ABCC12 orientation show a high similarity to those of ABCC4 and ABCC5 (Toyoda *et al.*, 2008; Yabuuchi *et al.*, 2001). No functional characterization has been reported so far for ABCC12 (Kruh *et al.*, 2007).

1.5.5.1.3. Breast Cancer Resistance Protein (BCRP, ABCG2 Subfamily)

ABCG2 subfamily is another ATP-binding cassette transmembrane transporter which transports a range of several drugs. It was first identified in MCF-7 human breast carcinoma cells, hence the name BCRP (Doyle *et al.,* 1998). Ross et al (1999) postulated that BCRP may be the main transporter that causes resistance to mitoxantrone in cancer cells (Ross *et al,* 1999). Exposure to mitoxantrone, topotecan, or doxorubicin results in over-expression of the ABCG gene in mice lacking P-gp and MRP hence the transporter is one of the three major transporters involved in multidrug resistance (Allen *et al.,* 1999; Doyle and Ross, 2003). BCRP also effluxes non-chemotherapeutic drugs and xenobiotics such as prazosin, glyburide, and 2-amino-1-methyl-6-phenylimidazo [4,5-b]pyridine (Ni *et al.,* 2010; Saito *et al.,* 2010). BCRP also mediates the intestinal efflux of antibiotics. For example nitrofurantoin which is an antibiotic used in treating urinary tract infection has a very high biliary excretion predominantly mediated by BCRP (Merino *et al.*, 2005b). Human BCRP and mouse bcrp1 can transport a range of organic substrates, including hydrophobic compounds, organic anions, weak bases, and conjugates of glucuronide, sulfate, glutamylate and glutathione of many endogenous and exogenous molecules. There is overlapping substrate specificity between BCRP and P-gp however the transport efficacies for these substrates differ (Ni *et al.,* 2010; van Herwaarden and Schinkel, 2006).

Tissue distribution of BCRP is similar to that of P-gp; BCRP is located in the apical membrane of epithelial cells of the intestines where it mediates direct intestinal excretion of its substrates and in the bile canalicular membrane of hepatocytes it stimulates hepatobiliary excretion (Allen *et al.,* 1999). Besides, BCRP has been shown to have protective role in blocking the absorption of drugs into CNS via the blood-brain barrier (Loscher and Potschka, 2005).

1.6. Assessment of drug-transporter Interactions

Transporters impact on both safety and efficacy in humans. Effect of transporter interactions on the therapeutic and other biological effects of drugs is complicated

due to the distribution pattern of these transporters in tissues and membrane localisations and varying, often complicated, roles in different tissue compositions. As a result, interaction of drugs with different transporters can impact their ADME properties and may lead to potential drug-drug interactions. In drug discovery, it is important to identify the possible drug-drug interactions for a drug candidate and evaluate the risk of occurrence in patient populations that are likely to receive a concomitant medication (Koenen *et al.*, 2011; Li, 2008).

Drug transporter interactions may be assessed using *in vitro* methods and they may be estimated using *in silico* techniques during drug discovery (Li, 2008). *Ex vivo* animal tissues have been traditionally used to measure drug permeability and transporter mechanisms, but since emergence of human overexpressing cell lines, these models have limited use in the industry (Obach *et al.*, 2012). The value of these assessments in drug development is to enable the prediction of drug-drug interaction risk in clinical settings (Li, 2008).

The experimental study of transporters requires the transporter expressed in a correct location of a plasma membrane (apical/basolateral) in correct orientation. During the experiment, the disappearance of drug substance from one compartment and/or appearance of the drug in the other compartment is/are measured. In order to measure the inhibition of a transporter by a drug, a validated specific substrate of that transporter is required to test the inhibitory activity against the transport of the substrate (Keogh, 2012). For example, hepatocytes can be grown in collagen sandwich cultures allowing them to establish the bile canaliculi necessary for directional flux to explore the impact of inhibitors on bile acid transporters (Kotani *et al.*, 2011, Maeda *et al.*, 2010, Marion *et al.*, 2011; Nakanishi *et al.*, 2011). In addition to primary hepatocytes, renal proximal tubule cells (Brown *et al.*, 2008) and brain microvessel endothelial cells (Lippmann *et al.*, 2012) are also used to mimic tissue barriers.

The experimental methods can generate quantitative or semi-quantitative measures such as binary data (substrate or non-substrate), IC_{50}, K_i, Km, Vmax, efflux ratio and intrinsic permeability. Michaelis-Menten model of enzyme kinetics are generally used to describe the interactions with transporters (Agnani *et al.*, 2011; Kolhatkar and Polli, 2010). Dissociation or association constant from the inhibitor-

enzyme complex and the concentration of the inhibitor to cause 50% inhibition at one chosen substrate concentration (IC_{50}) are some of the most common ways to present enzyme inhibition data (Li, 2008).

IC_{50} is defined as the required concentration of an inhibitor to inhibit the enzyme population by half (Copeland, 2005). IC_{50} can also be calculated from inhibitor concentrations and percentage of control activity using some non-linear regression methods (Chiba *et al.*, 2001). Typically, enzymes and IC_{50} determinations for the enzymes and transporters occur in early stage of preclinical development in order to generate preliminary inhibition data on a large set of compounds across a broad set of enzymes (Yan and Caldwell, 2001; Crespi and Stresser, 2000). However it must be noted that IC_{50} values can vary depending on the substrate used, the concentration of the labelled ligand (substrate) and different experimental variables and conditions (Böhm and Schneider, 2003). An advantage of IC_{50} determination is that it is independent of the inhibition mechanism and needs fewer samples to produce a meaningful result (Krishna, 2004). Nevertheless, the IC_{50} determination is dependent on the experimental and incubation conditions under which they are measured (Madan *et al.*, 2002). Thus, IC_{50} value is only meaningful at the substrate concentration for which the IC_{50} was determined for all forms of inhibition. Depending on the concentration of substrate used in the preliminary IC_{50} experiment, there can be a correlation between the IC_{50} and the inhibition constant (K_i) which can be used as an early approximation of K_i (Krishna, 2004).

Inhibition constant (K_i) plays an important role in predicting the clinical significance of inhibitions in *in vitro* methods. The K_i is a measure of enzyme-inhibitor potency and indicates how potent an inhibitor is. It is the concentration required to produce half maximum inhibition. In contract to the IC_{50} value, K_i is more reproducible because they are less dependent on experimental conditions as they are measured based on a range of substrate-inhibitor concentration (Krishna, 2004). IC_{50} value can be is converted to an absolute inhibition constant K_i by the Cheng-Prusoff equation. For enzymatic reactions, this equation is:

$$K_i = \frac{IC50}{1 + \frac{[S]}{Km}} \qquad \textbf{Eq. 1.12}$$

Where K_i presents dissociation constant of the inhibitor, [S] is fixed substrate concentration and K_m is the concentration of substrate at which enzyme activity is at half maximal (Cheng and Prusoff, 1973). In theory, a larger K_i value is an indication of low affinity and vice versa. For example in P-gp inhibition, the small K_i value means that substrate strongly blocked the P-gp and also means that enzyme-substrate complex (E-S) is more stable.

Although lab-to-lab variability is a well-established phenomenon for many experimental measurements, this may well be more pronounced for transporter assays using live cells, as many variables will impact on assay outputs including expression levels of the transporter, potentially endogenous transporters, passage number, assay formats (Keogh, 2012). A recent cross-pharma comparison of quantitative *in vitro* P-gp inhibition assays using a common substrate digoxin, with Caco-2, MDCK-MDR1 or P-gp vesicles, several assay end points, and data calculation methods showed limited agreement between assay outputs (Lee, 2011). The sources of variability are multi-factorial including cell-type, assay format and data manipulation (Bentz *et al.*, 2013).

For robust and reproducible *in vitro* transporter inhibition investigation, there is a need for characterised probe substrate(s) and inhibitors to determine the transport kinetic parameters such as initial rates, Km, Vmax, IC_{50} or K_i. In binary (yes or no) assays, there is a need for a single probe substrate concentration at or below Km, with and without inhibitors at concentrations sufficient to cause complete inhibition.

Although animals provide important *in vivo* mechanistic insights for transporters, their utility is limited, due to low throughput, the expense, and more importantly, the interspecies differences in transporter tissue distribution, expression levels and metabolism which limits the direct translation from preclinical species to humans (Obach *et al.*, 2012; Koegh, 2012).

1.7. *In silico* Methods in Drug Discovery

Traditionally, drugs are usually discovered in biological assays and in time-consuming *in vivo* and *in vitro* testing. However, the use of computer modelling in drug discovery has rapidly been developed creating techniques and software that are able to analyse and predict information about biological, chemical and medical data. The term '*in silico*' refers to the computational approach of drug discovery which is complementary to *in vivo* and *in vitro* experiments (Ekins *et al.*, 2007). In a widely expanding field, *in silico* techniques have been used to create virtual models that enable scientists to make predictions about biological activity and provide advances in medicine. Computational methods are used widely in drug discovery for the design of virtual compound libraries, identification of lead compounds (virtual screening), development of 3-D homology models for the biological targets, computing the interaction energies and geometries (protein-ligands docking), protein-protein interactions and estimations of biological activity of choice (Ekins *et al.*, 2002a). For example, quantitative structure-activity relationship (QSAR) has been applied for the analysis of growing collections of ADME data and the resulting models are used for the prediction of properties of new bioactive compounds (Golbraikh *et al.*, 2014).

In drug discovery, the use of computational methods to facilitate the discovery process is well established and plays an important role in modern drug discovery (Krogsgaard-Larsen *et al.*, 2010). Other commonly used *in silico* methods involve pharmacophore modelling that uses 3D structure representations to describe how candidate ligands may bind to a target (Ekins *et al.*, 2007). In addition, there are target based methods that include docking compounds to a target site and the use of scoring functions to score the binding affinity of the ligand to the target. It has gained popularity in recent times and has been involved in the discovery of inhibitors of HIV-1 integrase (Hayouka *et al.*, 2010).

1.7.1. Quantitative structure-activity Relationships (QSAR)

Since the 1960s when it was introduced by Corwin Hansch, QSAR has been used to describe the mathematical relationship between the structure of a molecule and biological activity (Van de Waterbeemd and Rose, 2003). QSAR models are empirical models in which a quantitative description of a chemical structure is related to the biological activity through an algorithm to guide future drug design (Cumming *et al.*, 2013). The predictive ability of QSAR models is directly influenced by dataset characteristics such as size and chemical diversity as well as employing different molecular modelling techniques, molecular descriptors, and statistical model development methods (Golbraikh *et al.*, 2014) and a thorough validation of the model for future predictions (Gramatica, 2013).

QSAR and other computer based methods can significantly reduce the time and the cost in drug design and discovery processes. Regression models in QSAR relate a set of predictor variables to the numerical potency of the response variable, while a classification algorithm relates the predictor variables to a categorical value of the response variable. The predictors consist of physicochemical and molecular properties of compounds and the QSAR response could be a biological activity of the compounds (Nantasenamat *et al.*, 2010).

The ability to predict a pharmacological activity is important. Predictive models are based on the given data, the technique to develop the model and the quality of information of the dataset. An ideal QSAR model should be simply understandable, interpretable and mechanistically relevant (Cronin *et al.*, 2010). A simple model should have a very small number of descriptors to form the relationship with the dependant. In QSAR, information and particular effect from molecular structure in a biological system can help us understand the relationship of molecular structure in a biological system (Cronin *et al.*, 2010).

1.7.1.1. Molecular Descriptors

The manipulation and analysis of chemical structural information is made possible through the use molecular descriptors (Leach and Gillet, 2003). According to Hong

et al. 2008, "Molecular descriptors are used to extract the structural information in the form of numerical or digital representation that is suitable for model development, serving as the bridge between the molecular structures and physicochemical properties or biological activities of chemicals". Molecular descriptors on a more mathematical based has been described by Todeschini and Consonni: "The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment" (Todeschini and Consonni, 2008).

Molecular descriptors play an essential role in chemistry and pharmaceutical sciences. Molecular descriptors are commonly used in QSAR for the identification and unique representation of molecules and fragments which are likely to become drug candidates (Malik *et al.,* 2006). Descriptors encode or map the structure of molecules into a set of numerical or binary values representing various molecular properties which explains activity (Dudek *et al.,* 2006).

Molecular descriptors are classified based on the compounds physiochemical property, topology, kappa shape indices, molecular finger prints, and pharmacophore keys (Dudek *et al.* 2006). The information contained in a molecular descriptor about a compound depends on the format in which the chemical is represented. This could either be a one-, two-, or three dimensional representations. One-dimensional (1D) descriptors represent mainly the molecular formula of the compound and describe only the bulk properties of the compound such as its molecular weight and number of specific atoms. Descriptors based on two-dimensional (2D) representations are able to provide information regarding atom types, connectivity patterns and topology such as number of aromatic group, number of hydrogen bond donors and acceptors, molecular refractivity, number of rotatable single bonds, bond distance and branching. 3D descriptors are more complex and provide information on conformation, geometry, potential energy such as dipole moment, ionisation potential, solvent accessible area, bond energy and solvation energy (Hong *et al.,* 2008).

1.7.1.1.1. 2D Molecular Descriptors

2D molecular descriptors are defined as numerical properties that can be calculated from the connectivity matrix, i.e. connection table representation, of a molecule but not from atomic coordinates. Therefore, the 2D descriptors are not dependent on the molecular conformation. As a result of this, they can be calculated quickly without the need for the optimisation of the three dimensional structures and are most suitable for large database studies. They can include physical properties such as sum of formal charges, bond counts, molecular connectivity and shape indexes (Hall and Kier, 2007), adjacency and distance matrix descriptors (Mihalic *et al.*, 1992), pharmacophore feature descriptors and partial charge descriptors.

Examples of 2D molecular descriptors provided by MOE software (Chemical Computing Group Inc. Montreal, Canada) include the van der Waals surface area calculated using a connection table approximation from 2D structure (vdw_area), octanol/water partition coefficient (log P), molecular mass density (density), sum of formal charges (Fcharge) and sum of the atomic polarisabilities (apol). The number of rings (rings), Lipinski's drug like test (Lipinski *et al.*, 2001) (lip_druglike), and number of aromatic bonds (b_ar) are examples of simple count descriptors, which are considered as 2D descriptors as they require 2D atomic connection map.

The Kier and Hall connectivity (chi, χ) and shape (kappa, κ) indexes are topological descriptors calculated from the hydrogen suppressed molecular graph (Hall and Kier, 1977; 2007). In addition, based on the same graph theory, the atom type electrotopological state indexes were suggested. These are atom level indexes that combine the electronic character of the atoms and the topological environment for each skeletal atom in a molecule (Kier and Hall, 1999).

Some 2D descriptors are calculated from adjacency or distance matrixes. The elements of an adjacency matrix for a molecule take the value of one if the two atoms are bonded and zero otherwise. The elements of a distance matrix of a chemical structure are the length of the shortest path between the two atoms. An example of descriptors calculated from adjacency matrix is BCUT descriptors (Pearlman and Smith, 1997). The BCUT descriptors are calculated from the eigenvalues of a modified adjacency matrix and are extensions of parameters

originally developed by Burden (1989). These parameters are based on a combination of the atomic feature for each atom and a description of the nominal bond-type for adjacent and nonadjacent atoms (Stanton, 1999).

Atomic partial charges can be combined by a variety of methods to calculate molecule level properties (descriptors). For example, total of all the negative atomic charges, or the sum of absolute charges can be calculated for a molecule to represent polarity of the molecule. In addition, van der Waals surface area of atoms with specific atomic charge ranges can be summed. An example of this is fractional positive van der Waals surface area (PEOE_VSA_FPOS) that can be calculated by MOE software (MOE Help file, 2012).

1.7.1.1.2. 3D Molecular Descriptors

3D descriptors are also known as shape-based descriptors as they depend on internal coordinates, conformation and three dimensional structure of the molecule. Such descriptors can be as simple as inter-atomic distances or torsion angles or as complex as the distribution of electrostatic potential around a molecule. Also similarity descriptors, allow comparison of the similarity of a molecule with a set of standard active molecules, on the bases of either electrostatic potential or steric parameters (Dearden and Cronin, 2005). An example of such molecular descriptors is dipole moment, which is controlled by the atomic charges, connection of atoms, and the three dimensional shape (internal coordinates) of the molecule. These computed 3D descriptors correlate well with the well-known experimentally observed physicochemical properties such as solubility (Kombo *et al.*, 2013).

Due the importance of the 3D shape, molecular structures need to be optimized (energy minimization) before the calculation of these descriptors (Akamatsu, 2002). Molecular orbital descriptors calculated by MOPAC are examples of these descriptors (Karelson *et al.*, 1996). Surface area, molar volume and shape descriptors and conformation dependent charge descriptors are other molecular descriptors that are dependent on the 3D shapes of molecules (Sauer and Schwarz, 2003).

Volsurf descriptors (known as the vsurf descriptors within the MOE program) were developed by Cruciani and co-workers (Cruciani *et al.*, 2000a) and noted as an important class of descriptors for the prediction of pharmacokinetic properties (Cruciani *et al.*, 2000b). These descriptors are calculated from 3D molecular fields of interaction energies also known as GRID (Goodford, 1985) molecular fields. In mathematical terms, these are 3D matrixes where the elements of the matrix are the attractive and repulsive forces between an interacting partner and a target. To calculate the Volsurf (and other molecular field) parameters, software first computes the fields by placing each molecule into a rectangular 3D grid (Leach and Gillet, 2003). Then a probe group is placed at each grid vertex and interaction energy between the probe and the molecule at points around the molecule is calculated (Goodford, 1985). For instance, MOE software calculates a parameter called vsurf_HB, which is calculated using a probe called O (carbonylic oxygen) to generate 3D H-bond donor fields (Fortuna *et al.*, 2008). The H-bond donor regions may be defined as the molecular envelope generating attractive H-bond donor interactions. H-bond donor descriptors can be calculated at different energy levels.

Other 3D molecular descriptors include electrostatic (E_ele) and van der Waals (E_vdw) components of the potential energy which can be calculated by semiempirical methods such as those implemented in the MOPAC engine in MOE software. The dipole moment (AM1_dipole), and the energy of the Highest occupied and the Lowest Unoccupied Molecular Orbitals (AM1_HOMO and AM1_LUMO respectively) are the examples of MOPAC descriptors that are calculated by AM1 semiempirical method (Stewart, 1993).

1.7.1.2. QSAR Model Development and Validation

QSAR models are statistically significant relationships between a biological property and molecular parameters of a set of compounds. The theoretical basis of classical QSAR is that the molecular structure is responsible for all the properties and biological activities of compounds and similar compounds should have similar biological and physicochemical properties (Katritzky *et al.*, 2001). Building a model that fits the available data is not adequate as the aim of any modelling

procedure is to be able to use the models for making future predictions. According to Gramatica (2011) 'an ideal QSAR should: 1) consider an adequate number of molecules for sufficient statistical representation, 2) have a wide range of quantified end-point potency (i.e. several orders of magnitude) for regression models or adequate distribution of molecules in each class (i.e. active and inactive) for classification models, 3) be applicable for reliable predictions of new chemicals (validation and applicability domain) and 4) allow to obtain mechanistic information on the modelled end-point.'

1.7.1.2.1. Statistical Modeling Techniques

A wide range of statistical techniques have been applied to the QSAR field. These can be classified based on the type of the data being modelled. Categorical data, such as the binary data types substrate/non-substrate or active/inactive, can be modelled using classification techniques that utilise the molecular descriptors in order to divide the data into the respective classes (Han and Kamber, 2006). Continuous data such as IC_{50} values can be subjected to prediction methods. Prediction methods, also known as regression-based methods, are used to predict missing or unavailable numerical data values rather than class labels (Han and Kamber, 2006). Among the regression-based approaches, the methods of multiple linear regression (MLR) and partial least squares (PLS) regression are prime examples in the QSAR field, while examples of classification methods involve, discriminant analysis and classification decision trees and support vector machines (Eriksson *et al.*, 2003).

Classification and prediction may need to be preceded by 'relevance analysis', which attempts to identify attributes that do not contribute to the classification or prediction process. These attributes can then be excluded. The commonly used terminology for this analysis in QSAR field is feature selection (Newby *et al.*, 2013a) or variable selection (Ghafourian and Cronin, 2006), or data reduction (Livingston, 2004). Due to the large numbers of molecular descriptors that are available through many commercially available software packages, variable selection has become a necessity in QSAR model development. This practice is essential to avoid overfitting to the training set data and the risk of chance

correlation (Ghafourian and Cronin, 2006). In addition, fewer molecular descriptors increase interpretability and understanding of resulting models (Weaver, 2004) and it can provide improved model performance for the prediction of new compounds (Norinder, 2003). Recently, 'descriptor pharmacophore' was introduced as a new concept in QSAR on the basis of variable selection. The descriptor pharmacophore is defined as a subset of molecular descriptors that lead to the most statistically significant QSAR models. It has been demonstrated that chemical similarity searches using descriptor pharmacophores as opposed to using all descriptors is more effective in successful mining of chemical databases or virtual libraries for identification of compounds with desired biological activity (Tropsha *et al.*, 1999; Tropsha and Zheng, 2001). Feature selection can be split into two broad categories: data pre-processing or embedded methods. Data pre-processing feature selection involves reduction of the number of molecular descriptors prior to incorporating them in the model development exercise. On the other hand, embedded methods incorporate the feature selection into the training of the model (Saeys *et al.*, 2007).

There are some unsupervised feature selection methods that do not use the dependent variable in the process of data reduction. An example of these methods, which can be used at pre-processing stage, is clustering of the variables. Cluster analysis is a useful tool for the visualisation of the clusters of variables as well as clusters of compounds (Livingstone, 2004). Another unsupervised method is Principle Component Analysis (PCA). This is multivariate technique in which a new set of variables called Principle Components (PCs) are created from linear combinations of original variables. PCs are orthogonal to each other and the first PC has the maximum information (variance) of the original data. Subsequent PCs describe the maximum of the remaining variance (Livingston, 2004). In this way, only the first few new variables (PCs) will be sufficient to explain the data and the remaining variables can be discarded, hence data reduction.

Other pre-processing techniques can be further split into filter and wrapper techniques. Filter techniques usually involve calculating a relative score of the molecular descriptors and ranking them in order of best score, and the descriptors that are at the top of the list are then used as input for classification. Wrapper techniques consider a number of subsets of molecular descriptors, evaluate each of

these based on the predictive performance of a classification model built from that descriptor subset, and eventually select the descriptor subset with the best predictive performance (Kohavi and John, 1997).

It's worth mentioning that any resulting QSAR model is only as statistically valid as the data that led to its development. In brief, an ideal QSAR model should consider enough number of training molecules, have a wide range of quantified endpoint potency for regression models and be applicable for prediction of new untested compounds (Gramatica, 2013).

To have a successful QSAR model, depends on accuracy of the input data and selection of appropriate descriptors should be considered (Chirico and Gramatica, 2012; Roy, 2007).

1.7.1.2.2. Validation of QSAR Models

The best fit models may not be the best ones for prediction. Only a stable and predictive model can be usefully interpreted for its mechanistic meaning, even though this is not always easy or feasible (Gramatica, 2011). The use of these statistical techniques in this context leads to 'statistical learning' from data that can be used for predictions. So far, much effort has been placed into performing some form of validation on QSAR models. Usually, this has been in terms of a model's statistical fit and more recently the focus has turned to using an external test set (Cronin, 2010).

Various strategies can be used for validation of QSAR models. According to Wold and Eriksson (1995) the most important validation strategies are: 1. internal validation set or a standard cross-validation method, 2. external validation by splitting the dataset into training set for model development and to evaluate the predictive ability of the model, 3. blind external validation (by using the model on a new external set), 4. data randomisation or Y-scrambling for verifying the absence of chance correlation between the dependent variable and descriptors (Wold and Eriksson, 1995).

The general idea of V-fold cross-validation is to divide the overall sample into a number of subgroups (V-folds). Subgroups are removed from the training set one at a time to serve as the internal test set and the model is developed successively for the remaining compounds (V – 1 folds). For each modelling run, some index of predictive validity is computed for the subgroup that is left out and the results of the v replications are averaged to yield a single measure of the stability of the respective model. The V-fold cross-validation technique is used in various analytical procedures to avoid overfitting of the data (Burden, 1989). V-fold cross validation is especially useful when the data is not large enough to allow for external validation of the model. The leave-one-out (LOO) method can be considered as a special case of V-fold cross validation. The outcome of this procedure is cross-validated R^2 (q^2), which is may regarded as a criterion of both robustness and predictive ability of the model. The robustness of LOO procedure has been debated recently (Kubinyi *et al.*, 1998; Golbraikh *et al.*, 2003).

Y-randomization is a widely used approach in validation of QSARs which is often used along with the cross-validation (Golbraikh *et al.*, 2003). It consists of repeating the model calculation procedure with randomized activities and subsequent probability assessment of the resultant statistics (Golbraikh *et al.*, 2003).

A more robust way for validation is to use external validation by splitting the dataset into training set, for model development, and validation set, to evaluate the predictive ability of the model. This is done before building the models so the validation set is kept external and not involved at any stage of model development. There are different methods for splitting the data into training and validation sets. It has been suggested that splitting data should be performed in a way that all representative compounds of the validation set are close to the training set compounds in the multidimensional descriptor space, and the representative points of the training set must be distributed within the whole area occupied by the entire dataset (Golbraikh and Tropsha, 2002.). The rational division of a dataset into training and test sets can be done by randomly allocating a fixed proportion of a homogeneous dataset to the validation set. In order for the training and validation sets covering similar activity ranges, the data could be ranked according to the

magnitude of the biological response, and every third or fourth chemical could be removed for validation set (Sharifi and Ghafourian, 2014). Other selection methods include selection on the basis of relevant physicochemical descriptors for example through multivariate design; this results in a test series of compounds in which all major structural and chemical properties are systematically varied at the same time (Eriksson *et al.*, 2003). An example of the other methods that can ensure similar distribution of training and validation set data is K-means-cluster based division of training and prediction sets (Leonard and Roy, 2008).

1.7.1.2.2.1. Applicability Domain

It is usually noted that QSAR is applicable only to compounds that are similar to the training set compounds (Katritzky *et al.*, 2001). Structurally limited training sets, when the dataset is small or when the chemical diversity is low, are a limitation of QSAR models in terms of their application for future predictions (Dimitrov *et al.*, 2005). A good model performance on the training set does not guarantee that a model will be predictive for validation set or external compounds (Stouch *et al.*, 2003). In other words, QSAR models sometimes are not applicable to the new compounds. As a result of this, there needs to be conditions set for the applicability of QSAR models (Eriksson *et al.*, 2003). This is very important in light of the increasing number of commonly termed global QSAR models which can be built on small datasets of low diversity (Weaver and Gleeson, 2008), or with poorly homogeneous training sets that contain partially overlapping clusters of compounds e.g. several classes of chemical compounds or chemotypes (Eriksson *et al.*, 2003). Defining a model's applicability domain is essential in order to determine the space of chemical structures that could be predicted reliably.

According to Weaver and Gleeson (2008) the domain of applicability is an important concept in quantitative structure-activity relationships (QSAR) that allows one to estimate the uncertainty in the prediction of a particular molecule based on how similar it is to the compounds used to build the model. In practice, there are various methods available for determining the range of applicability of QSAR models. For example, Dimitrov et al (2005) utilized a stepwise approach for

determining the applicability domain of QSAR models based on physicochemical properties in the training set of toxicity and skin sensitization datasets. This method involved four stages to account for the diversity and complexity of the QSAR models. First, the range of variation of the physicochemical properties of the training set compounds was specified. Then the structural similarities between chemicals that are correctly predicted by the model were assessed. At the third stage, the domain was defined based on a mechanistic understanding of the modelled phenomenon. Finally, the reliability of simulated metabolism was considered in assessing the reliability of predictions, if metabolic activation of chemicals is a part of the (Q)SAR model (Dimitrov *et al.*, 2005).

Sahigara et al (2012) has reviewed the applicability domain methods (Sahigara *et al.*, 2012). Accordingly, they have classified all the methods into: 1. range-based and geometric methods; 2. distance-based methods; 3. probability density distribution-based methods; 4. other approaches that may include decision trees and decision forests approach and stepwise approaches, such as the method suggested by Dimitrov et al (2005). Range based methods are the simplest approaches which may use a 'bounding box' defined on the basis of maximum and minimum values of each descriptor used to build the model or principle components of PCA (Netzeva *et al.*, 2005). In distance based methods, first the distance between an individual molecule will be computed from a defined point within the descriptor space of the training data using common distance measures e.g. Euclidean distance. Then, a threshold is applied to separate the compounds that are outside the domain of applicability. The threshold is a user defined parameter (Xu and Gao, 2003). As a distance based method, k nearest neighbour method can be used to measure the similarity by calculating the distance between the compound and the nearest neighbour compound in the training set (Xu and Gao, 2003). Probability density distribution-based methods are some of the most advanced approaches for defining applicability domain, as they are able to identify the internal empty regions within the data.

1.7.2. Enzyme-ligand Docking

Availability of a detailed 3D structure for biological drug targets (mainly receptors or enzymes) opens the possibility of a number of computer-based techniques in drug discovery arena. Structure-based drug design is one such technique that can use the information regarding shape and properties of the binding site of target molecules to design compounds which possess corresponding properties for fitting into and interacting with the binding site. Therefore, we require methods for determination of 3D structure of the biological targets (Krogsgaard-Larsen *et al.*, 2010). Other target based methods involve docking compounds to a target site and the use of scoring functions to score the binding affinity of the ligand to the target. It has gained popularity in recent times and has been involved in the discovery of inhibitors of HIV-1 integrase (Hayouka *et al.*, 2010) and aldose reductase inhibitors (Iwata *et al.,* 2001). Enzyme-ligand docking may guide a target's structural requirements for ligand (e.g. substrate/inhibitor) interaction by correlating the molecular features of validated ligands with their biological activity (Matsson *et al.*, 2007; Nicolle *et al.*, 2009; Ahlin *et al.*, 2008; Gombar *et al.*, 2004). The 3D structure of a protein can be obtained by prevalent methods such as X-ray crystallography and NMR spectroscopy, or predicted by homology modeling methods. The quality of an X-ray structure or a homology model is an important factor that should be taken into consideration before using the protein (Krogsgaard-Larsen *et al.*, 2010).

1.7.2.1. Conceptual Frame and Methodology of Molecular Docking

Computational approaches establish enzyme-ligand binding affinities by using structural information of the ligand and target enzyme, thus reducing the time and materials associated with experiments (Guvench and Mackerell, 2008). After X-ray crystallography or multidimensional NMR studies, the solved 3D structures of proteins are deposited into the Protein Data Bank (PDB) (RCSB Protein Data Bank, 2014). These structures can be analysed to discover the essential interactions and principles of molecular recognition (Raffa, 2001). The forces of interaction that bind a substrate to the enzyme active site consist of ionic bonds, hydrogen bonds,

van der Waals, hydrophobic, dipole-dipole and ion-dipole interactions. Once the interactions involved in substrate binding have been established, it is possible to look at the structure of a substrate and hypothesize the probable interaction that it will have with its active site (Schmidt *et al.*, 2013). The docking process involves the prediction of ligand conformation and orientation (or posing) within a targeted binding site. In general, there are two aims of docking studies: accurate structural modelling and correct prediction of activity (Kitchen *et al.*, 2004). Docking studies can be used to identify the fit between active site of the enzyme and the potential ligand. Also, docking can be used as a component of virtual screening, where a database of ligands is screened against a target protein (Kitchen *et al.*, 2004).

The docking process consists of two elements, the first is searches to find suitable conformation and the second is the measurement of the affinity of various conformations (Dror *et al.*, 2004). The process begins with the application of docking algorithms that positions small molecules in the active site. However, even relatively simple organic molecules can contain several conformational degrees of freedom. Conformational analysis is carried out to recognise conformational characteristic of ligand 3D structure created by energy minimization (Secundo, 2013). Energy minimization reduces the potential energy of a given conformation to make it suitable, but the obtained structure might not be essentially the most stable one as energy minimization stops when it reaches the first stable structure (the local minimum). To achieve the minimum with the lowest energy, structural variations will need to be carried out which helps in reaching the most stable conformation. In protein ligand docking, the docking program aims to find the preferred conformation of the ligand at a binding site of the target (Sousa *et al.*, 2006). Sampling of different conformations must be performed with sufficient accuracy to identify the conformation that best matches the receptor structure, and must be fast enough to permit the evaluation of thousands of compounds in a given docking run. The binding energy is then calculated for each conformation and is ranked and scored to give an estimation of the binding affinity between a compound and the target. Scoring functions are designed to predict the biological activity through the evaluation of interactions between compounds and potential targets (Kitchen *et al.*, 2004).

At present, there is a wide range of docking software available in the market with different scoring functions. The program AUTODOCK is one of the most cited docking programs and uses the Lamarckian genetic algorithm as well as a traditional genetic algorithm (Sousa *et al.*, 2006). GOLD is another program that is popular in the field and enables flexibility of the protein hydrogen bonds as well as the ligand being tested. Unlike AUTODOCK, docking scores in GOLD are ranked using a force field scoring function that includes the contributions of hydrophobic interactions, van der Waals forces and number of hydrogen bonds (Cummings *et al.*, 2005). FlexX is another software package that permits protein flexibility and scores the final position of molecules using the empirical Böhm's scoring function (Sousa *et al.*, 2006). In addition to these aforementioned programs, the Molecular Operating Environment (MOE) is a suite of applications that can be used for medicinal chemistry purposes. It includes a docking tool that searches for complimentary binding poses between a ligand and a rigid receptor which can be used to determine interactions between candidate ligands and targets.

1.7.2.2. Scoring Functions

Scoring functions are used to calculate the binding energy of poses generated after docking placements. A very accurate scoring function is desired to be able to successfully predict binding affinity, however due to the complexity and high computational cost involved, scoring functions make assumptions about molecular interactions based on experimental data from independent reactions (Lipkowitz and Boyd, 2002). In all scoring functions, a lower score indicates a more favourable pose while higher scores suggest that binding is less likely. Scoring functions are based on different calculation methods and can be divided into three categories: knowledge-based, force field and empirical based methods.

Knowledge-based functions use data from statistical analysis of structural complexes in the protein data bank, to estimate interatomic reactions occurring frequently between a ligand and the protein in specified intervals (Schulz-Gasch and Stahl, 2004).

GoldScore, Assisted Model Building and Energy Refinement (AMBER) and the Optimised Potentials for Liquid Simulations function (OPLS), are examples of force-field scoring functions. Force-field scores are calculated by measuring electrostatic and van der Waals interactions (Schulz-Gasch and Stahl, 2004) but are limited by the exclusion of solvation and entropic properties (Sousa *et al.*, 2006).
In contrast to these two scoring functions, empirical scores estimate free binding energy based on a sum of localised independent reactions (Lipkowitz and Boyd, 2002). In most cases, the constants in empirical formulas are derived from binding energies calculated in experiments of receptor-ligand complexes (Sousa *et al.*, 2006). An example of an empirical scoring function is the London dG scoring utilised in MOE (Equation 1.13).

$$\Delta G_{LdG} = c + E_{flex} + \sum_{h-bonds} c_{hb} f_{hb} + \sum_{metal-lig} c_m f_m + \sum_{atomsi} \Delta D_i$$

Equation 1.13. London dG Scoring Function (Corbeil *et al.*, 2012)

The formula above calculates binding energy, where E_{flex} represents the energy due to loss of flexibility of the ligand, f_{hb} and C_{hb} are measurements of hydrogen bonds, while C_M and f_M measure energies related to metal ligation.

Early scoring functions evaluated compound fits. Relatively simple scoring functions, on the basis of approximate shape and electrostatic complementarities, are heavily used during the early stages of docking simulations and in virtual screening of compounds. The selected conformers can then be further evaluated using more complex scoring schemes with more detailed treatment of electrostatic and van der Waals interactions, and inclusion of at least some solvation or entropic effects (Gohlke and Klebe, 2002).

2. Aims and Objectives

Biliary excretion is one of the main elimination routes of compounds and/or their metabolites with consequent effects on drug half-life and possible implications on gastro-hepatic cycle. The prediction of biliary excretion is a key target in the drug design and it helps with the selection of candidates for the development stage. The broad aim of the project involved not only the use of Quantitative Structure-Activity Relationships (QSAR) and data mining tools for estimation of biliary excretion, but also investigating the role of several transporter proteins in this elimination route. QSAR techniques employ molecular structural information of the compounds (molecular descriptors) and various statistical/ data mining techniques for the prediction of biological properties of new compounds. In this investigation QSAR methods were used to achieve both aims of the project.

Excretion of compounds through bile depends on the structural factors and physicochemical properties of the compounds. For example it has been suggested that compounds with molecular weights above 500 Da are highly excreted through bile (You and Morris, 2007) and this threshold value may be different in different animal species (Yang *et al.*, 2010). Therefore, it should be possible to link the biliary excretion fractions to molecular properties of drugs and obtain predictive tools using state-of-the-art data mining tools. At the first phase of the project, biliary excretion data measured as the percentage of intact compounds excreted through bile were collated from the literature. Several statistical analysis methods were used to develop and validate QSAR models to rationalize the effect of molecular structure of the compounds on the excretion of chemicals through bile. This is of a particular value during earlier stages of drug discovery where low-cost estimation procedures are required. Validation of each model would allow us to identify the most accurate models that could be used for the estimation of biliary excretion.

However, due to complexity of this disposition mechanism, that involves many transporter proteins for the active uptake of compounds from the blood into the hepatocytes and then active excretion of compounds from hepatocytes into the

canaliculus, previous QSAR models have encountered problems regarding the prediction accuracy when applied to new compounds (Gandhi and Morris, 2012). As a result, at the second phase of the project the binding of compounds to some of the important transporters with high expression levels in hepatocytes were used as an input parameter for the estimation of biliary excretion. The transporters that were considered at this stage included the efflux transporter P-gp and the uptake transporters OATP1B1, OATP1B3, and OATP2B1, which are known to have significant roles in biliary excretion of compounds (Pfeifer *et al.*, 2014; Kusuhara and Sugiyama, 2002).

To investigate the effect of binding to the above mentioned transporters on biliary excretion of compounds, it was necessary to investigate the structural requirements for binding. As a result, the objectives of this phase of the project included: 1. development of QSAR models for the binding of compounds to P-gp, OATP1B1, OATP1B3 and OATP2B1; 2. using the most accurate QSARs for each transporter to predict the binding activity of compounds in the biliary excretion dataset; and 3. incorporation of predicted transporter binding values in the QSAR models for the prediction of percentage excretion of compounds through bile. This workflow has been summarised in Figure 2.1.

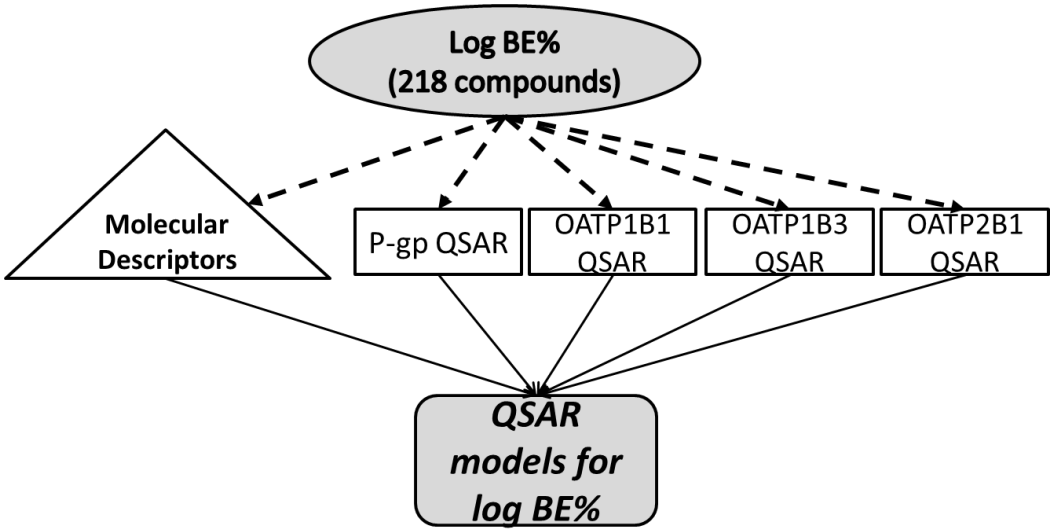

Figure 2.1. A diagram representing the phase II of this project

The application of computational modelling algorithms to obtain insight into transporter-substrate interactions has met with increasing success by the availability of high-quality datasets and atomic resolution structures of some transporters (Giacomini *et al.*, 2010). A further objective of this project was to investigate ligand-transporter docking as a prediction tool for the estimation of binding of compounds to the transporters. The 3D structure of mouse P-gp was available through Protein Data Bank (PDB) website and the substrate binding site has been proposed recently by (Aller *et al.*, 2009). The score of docking experiment was used as a molecular descriptor for the prediction of compounds binding to P-gp.

There are a number of linear and non-linear prediction (regression-based) methods and classification methods that are available for the statistical model development in QSAR studies. Each method may offer a number of advantages and disadvantages. In this investigation, the aim was to use a combination of various available methods in order to achieve the best predictive models. The methods included stepwise regression analysis, Classification and Regression Trees (C&RT), Chi-square Automatic Interaction Detector (CHAID), Boosted trees (BT), Random Forest (RF) and Multivariate Adaptive Regression Splines (MARS) models.

QSAR models developed in this study for the prediction of biliary excretion can be very useful in explanation and understanding of the clearance system as well as selecting the candidate drugs in selection processes in drug discovery. Moreover, the models resulted in some insight into major factors that can affect biliary elimination of drugs.

3. Methods

The major methods employed in this work consisted of various QSAR and molecular docking techniques that were used for the estimation of biliary excretion and binding of compounds to the transporters, P-gp, OATP1B1, OATP1B3 and OATP2B1.

3.1. Datasets

The datasets for each investigation have been explained in the relevant chapters (Chapters 4-6). Table 3.1 gives a summary of the datasets.

Table 3.1. Summary of the datasets used.

Dataset	N	Data type
Biliary Excretion	217	Percentage of intact dose excreted through bile in rats (log BE%)
P-gp binding	219	Inhibition constant (log K_i) measured *in vitro*
OATP binding	225	Percentage inhibition measured *in vitro*

3.2. Calculation of Molecular Descriptors

3.2.1. ACD Labs/LogD Suite 12.0.

Simplified Molecular Input Line Entry System (SMILES) notations for all compounds were obtained by search in systematic names in ACD/dictionary (ACD Labs/LogD suite version 12.0., Advanced Chemistry Development Inc., Ontario, Canada). If compounds were not available in the ACD/dictionary, then ChemFinder gateway version 3.0 (CambridgeSoft, USA) was utilized to obtain the molecular structure. Moreover, SMIELS codes were double-checked in the online database ChemSpider (approved by the community of Royal Society of Chemistry - RSC) (ChemSpider, 2001). The SMILES notation of each compound was generated either by entering the systematic name of the compound in the ACD/Dictionary to acquire their molecular structures and SMIELS codes or by drawing the structure in the software and then obtaining the SMILES for the drawn structure.

Compound names and SMILES codes from Excel were copied into a Notepad file, and saved in txt format. Notepad file was imported into ACD history view and different physicochemical properties were calculated for all compounds. The properties included logarithm of the octanol/water partition coefficient (LogP), logarithm of apparent partition coefficient (LogD) at different pH values 2, 5.5, 6.5, 7.4 and 10, dissociation constant (pKa) for acidic and basic compounds, molar volume, index of refraction, polarisability, polar surface area and others.

Fraction of compounds ionised at pH 7.4 were calculated from dissociation constants (pKa). The fractions of compounds that is ionised at pH 7.4 as acid (FiA), as base (FiB), or (for zwitterionic compounds) as acid and base (FiAB), and the fraction unionised (Fu) were calculated from the lowest acidic and the highest basic pKa values and are presented in equations 3.1 to 3.4 respectively (Ghafourian *et al.*, 2006).

$$\text{FiA} = \frac{1}{1+\text{antilog}(\text{pKa}-7.4)} \qquad \textbf{(Eq. 3.1)}$$

$$\text{FiB} = \frac{1}{1+\text{antilog}(7.4-\text{pKa})} \qquad \textbf{(Eq. 3.2)}$$

$$\text{FiAB} = \text{FiA} \times \text{FiB} \qquad \textbf{(Eq. 3.3)}$$

$$\text{fU} = (1-\text{FiA}) \times (1-\text{FiB}) \qquad \textbf{(Eq. 3.4)}$$

In Equations 3.1 and 3.2, pKa is the most acidic and the most basic pKa, respectively, which were obtained from ACD Labs pKa database and, in case the experimental pKa was not available, it was calculated by the software.

The ACD/LogD calculations were performed for all compounds and the results were transferred to Microsoft Excel worksheet.

3.2.2. TSAR 3D

Using TSAR 3D software (Version 3.3., Accelrys Ltd.) additional molecular descriptors were calculated. The SD file created by ACD software was imported into TSAR 3D. In this software, each row stored information about one compound and each column stored a molecular descriptor. Initially, the partial atomic charges were calculated for the molecules and COSMIC optimize 3D was applied to minimize the molecular potential energies. This was essential since the generation of 3D descriptors needs to be based on an accurate 3D molecular structure and geometry. However, due to errors in some of the imported structures, COSMIC energy minimisation did not automatically work for some of the compounds. Hence the 3D structures of these compounds were modified manually by using the 3D visualise tab in TSAR 3D to correct the errors and then run the COSMIC minimisation. In most cases the structural errors were due to the valence state of atoms which varied between ACD generated SD files and those in TSAR 3D. For some of the compounds the SD molecular file format could not be used and the SMILES codes were imported to TSAR instead. The SMILES codes and the compound names were copied and pasted in MS-Word as 'text'. Using the "Find" icon, the document was edited by finding "Tab" and replacing with "space". The edited document was then copied into WordPad and saved as text with .smi file extension. The codes were then imported into TSAR 3D and eventually cosmic minimisation was successfully executed. In few cases, calculations by TSAR 3D were not possible. For example, the presence of heavy metal Platinum (Pt) in the structure of a compound would lead to such an error.

A series of descriptors consisting of electronic, steric and hydrophobic parameters as well as topological indexes were calculated using TSAR 3D for each compound. The quantum mechanical properties were calculated using VAMP electrostatic routine in TSAR 3D. The method used in VAMP was the semi-empirical approach, AMI Hamiltonian. The calculated quantum mechanical properties include electronic energy, total energy, accessible surface area, mean polarisability, dipole moment, energy of the highest occupied molecular orbital (HOMO), and energy of the lowest unoccupied molecular orbital (LUMO). VAMP calculations were not possible for compounds with more than 50 heavy atoms in their molecular

structures. The minimized molecular structures were saved as a SD file and the molecular descriptors were exported to Excel.

3.2.3. Molecular Operating Environment (MOE)

The saved SMILES codes and names from ACD Labs/LogD were imported into MOE software (Version 2012.10, Chemical Computing Group Inc. Montreal, Canada). Using the wash tab, any unwanted fragments including salts and water molecules were removed from the molecular structures. This process also neutralized the protonated state of any charged structure.

Following the wash procedure, energy minimization was carried in order to calculate atomic coordinates corresponding to the local minima. Within the energy minimization function, the "preserve existing chirality" was also selected. Thereafter, self-consistent field (SCF) calculations were performed. The SCF energy minimization technique constructs an initial guess density matrix, in terms of the atomic orbitals and then iteratively refines them by correcting the kinetic energy, nuclear energy and electron – electron repulsion. This allows the density matrix to be self-consistent. The parameters calculated by SCF for the minimized structures were SCF energy, HOMO-LUMO energy gap, heat of formation and dipole moment.

Finally, after SCF energy minimization, all molecular descriptors were calculated for each of the compounds and all data were saved as SD format and exported to Excel.

3.2.4. Symyx QSAR version 2.2

Symyx QSAR software (previously known as MDL-QSAR) was used to obtain additional molecular descriptors for the compounds in the datasets. Symyx QSAR can calculate many new molecular descriptors such as atom type electrotopological indexes. The SD file from MOE was imported and electrotopological state indexes

for different atom types along with other topological indexes were calculated. The molecular descriptors were then exported into an Excel file.

3.3. Development and Validation of QSAR Models

In this work, various data analytical techniques were used for the development of QSAR models. Datasets of compounds were first divided into training and external validation sets. In order for the training and test sets to have a similar range of biological activities, compounds in each dataset were ordered according to the relevant response variable and, depending on the size of the dataset, from each group of five or four compounds one was allocated into the external validation set. Models were developed using the training set compounds. These models were used for the estimation of the response variable for the external validation set. The details of these processes for individual datasets have been explained in the relevant chapters.

Goodness-of-fit in prediction (regression based) models

Discrepancy between observed and predicted values shows the error, and is used to assess the accuracy of QSAR models. The mean absolute error (MAE), root mean squared error (RMSE) and mean fold error (MFE) were utilised to assess the accuracy of predictions by QSAR models.

$$\text{MAE} = \frac{\sum |(\text{observed} - \text{predicted})|}{\text{N}} \qquad \textbf{(Eq. 3.5)}$$

$$\text{RMSE} = \sqrt{\frac{\sum (\text{observed} - \text{predictated})^2}{\text{N}}} \qquad \textbf{(Eq. 3.6)}$$

$$\text{Mean Fold Error} = \text{antilog}\left(\frac{\sum |\text{log.observed} - \text{log.predicted}|}{\text{N}}\right) \qquad \textbf{(Eq. 3.7)}$$

Calculation of error in classification models

There are extensive numbers of performance measures used to validate the predictive power of classification models. The performance of each algorithm was measured using three performance measures, sensitivity (SE), specificity (SP) and overall accuracy.

Sensitivity is proportion of compounds correctly predicted to be positive relative to all the compounds experimentally determined to be positive:

$$\text{Sensitivity} = TP/ (TP + FN) \quad \textbf{(Eq. 3.8)}$$

Where TP is number of true positives, TN is number of true negatives, FP is number of false positives, and FN is number of false negatives.

Specificity is proportion of compounds correctly predicted to be negative relative to all the compounds experimentally determined to be negative:

$$\text{Specificity} = TN/ (TN + FP) \quad \textbf{(Eq. 3.9)}$$

Overall accuracy in this study is defined as:

$$\text{Overall Accuracy} = SP \times SE \quad \textbf{(Eq. 3.10)}$$

3.3.1. Stepwise Regression Analysis

Minitab Statistical Software Version 16 was used for the development of multiple linear regression (MLR) models. In stepwise regression analysis, independent variables were normally all the molecular descriptors and the dependent (response) variable was the activity under investigation. For example logarithm of the percentage dose excreted via bile (log BE%) was the dependent variable in Chapter 4. In all regression analyses, a P value of less than 0.05 was considered to be statistically significant for the variables. Values for “alpha to enter” and “alpha to remove” items in stepwise regression method were set to 0.05.

3.3.2. Classification and Regression Trees (C&RT)

Introduced by Breiman in 1984, C&RT are decision tree algorithms that produce classification or regression trees depending on whether the dependent variable is categorical or numerical. The analysis uses the Gini coefficient as an identifier of suitable splitting criteria (Breiman *et al.*, 1984). Based on recursive partitioning, C&RTs are constructed by successively splitting a dataset into increasingly homogeneous subsets until it is infeasible to continue, based on a set of stopping rules (StatSoft, 2009). The analysis has an embedded feature selection method which picks the most significant molecular descriptors for splitting the data into the two most homogeneous groups (called branches or nodes).The process works by monitoring the error on the test data during growth and choosing the one with minimal error (Breiman *et al.*, 1984). This algorithm starts off with the complete training set, evaluates all available attributes (e.g. molecular descriptors), choosing the one which best separates it. It then recursively proceeds to split the resulting subsets until no improvement can be made by continuing to split; this happens when the tree reaches a certain complexity based on the pre-set stopping criteria or until all the data in the nodes have the same value.

STATISTICA software has Classification and Regression Trees (C&RT) routine, which can develop classification tree (CT) or regression tree (RT) by selecting the most significant molecular descriptors out of the descriptor pool at every step of partitioning. C&RTs can also be built interactively, using the manually selected descriptors.

Stopping rules are the criteria used to find the right-sized tree. The size of a tree in C&RT analysis is an important issue, since an unreasonably big tree can lead to overfitting and make the interpretation of results more difficult. Stopping parameters could be a combination of the minimum number of cases, the maximum number of levels, the maximum number of nodes, and minimum fraction of objects for splitting. The parameters have mainly to do with which nodes should be split and which should be terminal nodes. STATISTICA offers two choices for stopping nodes: 1. Prune variance, and, 2. FACT direct stopping. When using deviance, the minimum number of cases and maximum number of nodes are used for stopping. For example with minimum number of cases equal 100, a node with less than 100

cases will be a terminal node and no further split will be made. The maximum number of nodes controls the overall tree complexity. The default stopping parameters in STATISTICA software depend on the number of data points (number of compounds). For the FACT style stopping method, fraction parameters, rather than number of compounds, will determine if a node should be split.

The advantage of C&RTs is their simplicity at interpretation of results summarized in a tree. The final results of using tree methods for classification or regression can be summarized in a series of logical if-then conditions (tree nodes). Therefore, there is no implicit assumption that the underlying relationships between the predictor variables and the dependent variable are linear.

3.3.3. Interactive Tree (I-tree) Using C&RT

Interactive tree is a C&RT-style tree, which allows for the molecular descriptors to be selected manually by the operator. This tool is useful when investigating the effect of certain variables/ molecular descriptors on the property under investigation. In I-tree, apart from the usual V-fold cross-validation procedure, another cross-validation option, "Cross-validate tree sequence" was also applied. This validation method is applied to the entire tree sequence, instead of just the final tree in V-fold cross-validation (Hill and Lewicki, 2006).

3.3.4. Chi-square Automatic Interaction Detector (CHAID)

The Chi-square Automatic Interaction Detector (CHAID) is one of the oldest decision tree methods initially suggested by Kass in 1980 (Kass, 1980). This tool performs multi-level splits where C&RT uses binary splits. CHAID is well suited for large datasets. Cross validation either V-fold or train and test samples can be used to safeguard against overfitting the CHAID tree. The Stopping criteria includes minimum number of cases for splitting, maximum number of nodes, probability for splitting and probability for merging. To test the statistical significance of splits, CHAID computes a Bonferroni adjusted P-value for the

respective descriptor (Hill and Lewicki, 2006). Bonferroni adjustment is an option in CHAID, used to control the type one error rate (familywise error rate) when testing multiple hypotheses. It usually is accomplished by dividing the alpha level by the number of tests being performed (usually 0.05 / n). In this work, we employed Bonferroni adjustment as our preliminary results showed lower cross validation error when this adjustment was used.

3.3.5. Boosted Trees (BT)

Boosted trees analysis generates a series of very simple boosting regression trees (BT) where each successive tree is built for the prediction of residuals of the preceding tree. Each of these trees has a weak predictive accuracy, but using the weak predictors together can create a strong predictor (Lewicki and Hill, 2006). The user defined parameters in this analysis includes the learning rate, the number of additive terms (number of trees), random test data proportion (fraction of data points in testing pool) and subsample proportion The seed for random number generation that controls which cases are selected in sampling was set to one. The maximum number of nodes was set to three, which means that each tree will have just one binary split.

3.3.6. Random Forest Trees Model (RF)

A random forest (RF) model is an ensemble of tree predictors such that each tree depends on the values of a random vector (a random selection of molecular descriptors and training set compounds) sampled independently. The method builds a series of simple trees where the predictions are taken to be the average of the predictions of all the trees (Breiman, 2001). The analysis removes a user defined portion of the data and keeps it as the internal test data. The remaining training set data is sampled consecutively and models are developed for each subsample. Various subsample proportions along with different numbers of trees may be selected. The number of predictors (to be randomly considered at each node) was set to nine throughout the thesis. The default settings were used for stopping

conditions including minimum number of cases, maximum number of levels, minimum number in child node and the maximum number of nodes which is different depending on the size of the dataset. The best model was selected based on the estimation error for the internal test data.

3.3.7. Multivariate Adaptive Regression Splines (MARS) Model

MARS is a non-parametric regression procedure that constructs a relation between the dependent and independent variables from a set of coefficients and basic functions that are entirely driven from the regression data (Friedman, 1991). It is a very flexible technique that automatically models non-linearities and interactions between variables. The non-linearities (knots) are represented by the so called 'hinge functions' which are expressions of the type 'max (a, b)' where the value of this expression will be a if a>b, or else b. Interactions between each variable pairs can also be expressed in the formula. MARS model is developed by stepwise addition of basis function in pairs (forward pass) to reduce the sum-of-squares residual error, and then step-by-step removal of the least significant terms to achieve better generalisation (backward pass). The model created by this tool is easy to understand, compared to some other data mining models such as boosted trees. This tool sometimes is used as a method for finding the important predictor variables as important information for another analysis. The MAR Splines algorithm picks up only those basis functions (and those predictor variables) that makes a "sizeable" contribution to the prediction. Basis functions use a non-parameter (break point) to find non-linear relationships. Increasing the maximum number of basis functions gives the potential for more complex model. Using the degree of interaction we can specify no interaction up to a very high order interaction term. Model subsets are compared using the GCV criterion (Generalized Cross-Validation). GCV is the adjusted form of residual sum-of-squares that penalises the addition of knots in order to limit the model flexibility and overfitting.

In this investigation, in addition to using all the molecular descriptors in MARS analysis and allowing MARS to select the significant descriptors, we performed a pre-processing feature selection to select a limited number of molecular descriptors

for use in MARS analysis. The feature selection methods were different for different datasets and have been explained in the relevant chapters.

4. QSAR Models for Biliary Excretion

4.1. Introduction

Biliary excretion is an important route for the elimination of some drugs and their metabolites (Rosenbaum, 2011). Although the liver is generally identified with its role in metabolism, one of the most important functions of the liver is formation of bile which is then stored in the gallbladder and discharged into the duodenum upon ingestion of food, with bile carrying also cholephilic xenobiotics. Bile which is a composition of bile acids and other components such as phospholipids, bilirubin and cholesterol is formed in the hepatocytes and is actively discharged across the canalicular membrane into canaliculus (Rollins and Klaassen, 1979). Once bile is released into the intestine, some metabolites and unchanged drugs continue their way of elimination through the faeces. Others, for example lipid-soluble drugs, are reabsorbed from the intestine and move to the systemic circulation (Rollins and Klaassen, 1979). This enterohepatic circulation affects pharmacokinetics by keeping the plasma concentration of drugs high (Rosenbaum, 2011). Enterohepatic cycling and biliary elimination can continue until the compound is ultimately eliminated from the body by faecal or renal excretion or metabolism. Uptake from sinusoidal blood and then secretion of bile salts across the canalicular hepatocyte membrane are the major factors controlling the rate of bile secretion.

Basolateral bile salt uptake is driven through the Na^+-dependent and Na^+-independent uptake systems (Kullak-Ublick *et al.*, 2000). The main sodium-dependent bile salt transporters are Na^+-taurocholate co-transporting polypeptides (human and rat). On the other hand, the Na^+-independent uptake of bile salts cannot be attributed to the function of a single transport system and several carrier systems have been implicated including sulphate/anion exchanger, dicarboxylate/anion exchanger and OH^-/cholate exchanger. In rats, the organic anion transporting polypeptides (Oatp1, Oatp2 and Oatp4) have been indicated as the main sodium-independent uptake proteins (Kullak-Ublick *et al.*, 2000). The organic cation and organic anion transporters (OCT and OAT, respectively) also play important roles in the initial sinusoidal influx of drugs into hepatocytes (van Montfroot *et al.*,

2003). These transporters have wide substrate specificities for a range of exogenous and endogenous substrates (Leabman *et al.*, 2003). OCT1 can be found abundantly in hepatocytes and may be seen as the most important transporter for distribution of cationic compounds into the liver from sinusoidal membrane (Nies *et al.*, 2009).

Canalicular bile secretion is an osmotic process in which active excretion of organic solutes into the bile canaliculus is the main driving force for the passive inflow of water, electrolytes and nonelectrolytes from hepatocytes (Trauner and Boyer, 2003). While products of the multidrug resistance gene family (Mdr), namely bile salt export pumps, Bsep (rat) and BSEP (human), transport monovalent bile salts (Rollins and Klaassen, 1979), excretion of non-bile salt organic anions and divalent sulphate or glucuronide bile salts is carried mainly by the multidrug resistance protein 2 (MRP2). Bile salt export pump has a limited role in drug excretion. However, drug inhibition of this pump can lead to hepatotoxicity (Morgan *et al.*, 2010). Another member of this family, P-glycoprotein, also has known as multidrug resistance protein 1, actively effluxes xenobiotics into the bile (Schinkek *et al.*, 1997). Breast cancer resistance protein (BCRP/ABCG2) is also involved in the transport of a range of drugs. For example, nitrofurantoin has a very high biliary excretion predominantly mediated by BCRP (Merino *et al.*, 2005b). Other basolateral isoforms of the multidrug resistance-associated protein, MRP4 and MRP3, provide alternative routes for the elimination of organic anions from hepatocytes into the systemic circulation (Kullak-Ublick *et al.*, 2000). Properties of the chemical structure as well as the characteristics of the liver such as specific active transport sites within the liver cell membranes are the main factors which determine the elimination of xenobiotics via the biliary tract (Rollins and Klaassen, 1979). Despite the various transport systems involved in the biliary elimination of xenobiotics, there has been a number of attempts to identify common molecular features of highly excreted compounds. Molecular weight (MW) has been suggested as an important factor in biliary excretion levels of compounds. Anionic compounds with the MW higher than 325±50 kDa in rats, 400±50 kDa in guinea pigs, 475±50 kDa in rabbits and 500±50 kDa in human have been suggested as good candidates for biliary excretion (Hirom *et al.*, 1972). Most compounds with lower molecular weights are quickly cleared through the kidneys and are not excreted in the bile (Abou-El-Makarem *et al.*, 1967). Bile is rich in endogenous

organic anionic substrates (e.g., steroid hormones), organic cations (such as quaternary ammonium), bilirubin and bile acids (Rollins and Klaassen, 1979). Moreover, excretion route of anionic xenobiotics and some antibacterials is through the bile (Crosignani, 1996). Principally, for organic cationic compounds, biliary elimination depends on the molecular volume (Neef *et al.*, 1984), lipophilicity of the compound and the number of cationic groups (Feitsma, 1989).

Biliary excretion has major significance in determining the pharmacokinetic profiles of drugs. In several disease states, the excretion of drugs through bile is affected and toxicities may arise (Rosenbaum, 2011; Rollins and Klaassen, 1979). Knowledge of biliary excretion levels of compounds can help in identifying any possible mechanisms of hepatobiliary toxicity and potential drug-drug interactions. Therefore, an insight into the structural profile of cholephilic compounds through accurate modelling of the biliary excretion is important for predicting clinical pharmacokinetics. This is of a particular value during earlier stages of drug discovery where low-cost estimation procedures are required. Quantitative structure-activity relationships (QSARs) employ data mining techniques to explore the relationships between biological properties of interest, e.g. pharmacokinetic parameters of drugs, and the properties of the molecular structures (Ghafourian *et al.*, 2006). Recently, a QSAR model developed using 2D molecular descriptors showed good prediction ability for a set of literature biliary excretion data measured under the same experimental model (Luo *et al.*, 2010). However, re-evaluation of this simple model showed that the statistical significance of the model is lost when it is used for the prediction of a wide set of external compounds (Gandhi and Morris, 2012), suggesting that hepatobiliary excretion cannot be captured by simple physicochemical descriptors when examining chemically dissimilar compounds. Unfortunately, availability of *in vivo* biliary excretion data which is necessary for modelling is very limited. Yang *et al.* (Yang *et al.*, 2009) have recently compiled a big dataset of percentage of dose eliminated in the bile in rats and humans. This offers an excellent resource for a detailed study on the structural determinants for high biliary excretion. Using this dataset, Yang and co-workers suggested a MW threshold of 400 Da for anions in rats and 475 Da for anions in humans. They also developed linear regression models for human and rat. The aim of this study was to use an expanded dataset and incorporate non-linear methods to develop statistically

valid QSAR models. Specifically, classification and regression tree (C&RT) is a flexible and yet simple and interpretable technique with embedded feature selection that selects the most significant molecular descriptor for partitioning the data into smaller subsets of similar observations (Breiman *et al.*, 1984). This rule-based technique is a decision tree that splits the data in a recursive manner until the subset has all the same value of the target (dependent) variable, or when no gain in the prediction accuracy is achievable by further splitting. In this study, we aimed at using regression trees and two ensemble methods that construct many such decision trees and return the consensus prediction by the trees, namely random forest and boosted trees. The prediction accuracy of the models and the molecular descriptors selected by these methods were compared in order to clarify the structural elements controlling the biliary excretion. Moreover, regression trees were used to examine the significance of molecular weight and presence of carboxylic acid groups and to find the statistically significant threshold values. In this case, regression trees are useful since they can be used interactively so that a molecular descriptor of choice can be incorporated at any split level and the analysis may determine the statistically significant threshold value of the descriptor for splitting the data.

4.2. Methods

In this investigation RT models were made with log BE% as the dependent variable and predictors were selected by this statistical analysis from all the molecular descriptors used in the analysis. “observed” refers to the log percentage of intact dose excreted into the bile from *in vivo* studies. In all statistical analyses, logarithm of percentage dose excreted (log BE%) was used in the analysis instead of percentage of dose excretion. This was due to the normal distribution of log BE% as indicated by the skewness comparison with BE.

4.2.1. The Dataset

The biliary excretion dataset was that collated by Yang et al (2010) available at http://www.buffalo.edu/~memorris, with the addition of some new data from

literature (Hirom *et al.*, 1972., Abou-el-makarem *et al.*, 1967., Hughes *et al.*, 1973., Fahrig *et al.*, 1989., Funakoshi *et al.*, 2005., Luo *et al.*, 2010., Scott *et al.*, 1994., Matsushita *et al.*, 1992., Prueksaritanont *et al.*, 2003., Niinuma *et al.*, 1999., Vaidyanathan *and* Boroujerdi, 2000., Fukuda *et al.*, 2008., Chu *et al.*, 1997., Wu *et al.*, 2008., Wright and Line, 1980., Chan *et al.*, 2002., O'Reilly *et al.*, 1971., Watkins Dykstra, 1987., Sasabe *et al.*, 1999., Weinz *et al.*, 2009., Mohri *et al.*, 2005., Kemmerer *et al.*, 1979., Itagaki *et al.*, 2003., Evanchik *et al.*, 2009., Krishna *et al.*, 1999., Broggini *et al.*, 1980., Israel *et al.*, 1978., Arimori *et al.*, 2003., Itoh *et al.*, 2004). It consists of *in vivo* biliary excretion expressed as percentage of dose excreted as the parent compound intact through the bile (BE%) for 217 compounds in rat after iv or intraperitoneal administration of the compound. The compounds are from different chemical classes such as bile acids, statins, dyes, penicillins and cephalosporines, macrolide antibiotics, quinolone antibiotics, NSAIDs, thrombin inhibitors, analgesics, anti-cancer drugs such as doxorubicin, folate, peptides, anti-HIV agents, quaternary ammoniums, sulphanilamide and arylaminosulphonic acids. Biliary excretion in the database is presented by percentage of drugs excreted through bile, or bile clearance.

Where several values were available for the same compound the mean values were used. Table 4.1 shows an example of this.

Table 4.1. Example of different values for the same compound.

Compound	% Dose excreted in bile as parent compound	Collection period
Methotrexate	72	480 min (8 hr)
Methotrexate	84.3	600 min (10 hr)
Methotrexate	58.9	720 min (12 hr)
Methotrexate	64	1440 min (24 hr)
Average	69.8	

This dataset is presented in Appendix I, including all the references.

4.2.2. Model Development and Validation

In this study, QSARs were established to relate the biliary excretion of compounds (log BE%) to the molecular descriptors. Molecular descriptors were calculated according to the procedures explained in section 3.1. Before building the models, the molecular descriptors were checked to find and discard those columns containing more than 98% constant values or more than 28 (out of 217) missing values. The total number of molecular descriptors used in all statistical analyses was 387.

The compounds were divided into an external validation set and a training data. To divide the compounds, they were ordered according to BE%, and from every set of five compounds, four were allocated into the training and one into the external validation set randomly. In this way, training data consisted of 168 compounds and the external validation set consisted of 40 compounds. For the analytical methods that required parameter optimization, a fraction of training set compounds were randomly assigned into internal validation set, or alternatively, cross validation was used if the option was available in the statistical software. STATISTICA Data Miner was the software used for statistical analysis. The general idea of V-fold cross-validation is to divide the overall sample into a number of V-folds. The V-fold cross-validation technique is used in various analytical procedures to avoid overfitting of the data (Burden, 1989). For the internal validation set, where applicable, the risk estimate and standard error were calculated in STATISTICA software and used as the performance indicators. Risk estimate is calculated as the proportion of residual variance incorrectly estimated by the model. Standard error measures the error of the prediction.

Several linear and non-linear methods were used for the QSAR model development. These included stepwise regression analysis, stepwise regression analysis, Classification and Regression Trees (C&RT), Boosted trees (BT) and Random Forest (RF).

The methods have been explained in section 3.2. In C&RT analysis, several stopping criteria were examined, including the default settings in STATISTICA. The default stopping criteria were minimum number of cases of 21 and the

maximum number of nodes set to 100. The default V-value of 10 was used in the V-fold cross-validation and the risk-estimate was used to check the reliability of the resulting RTs. In BT analysis, the default values for learning rate, the number of additive terms, random test data proportion and subsample proportion were 0.1, 200, 0.2 and 0.5 respectively. Various subsample proportions of 0.4, 0.45, 0.50, 0.55 and 0.60 were examined in combination with the learning rates of 0.1 and 0.05. The best two models were selected based on the performance indicators for the internal validation set. In RF analysis, various subsample proportions of 0.40, 0.45, 0.50, 0.55 and 0.60 were examined. Different numbers of trees were tested at 20, 50, 80, 100 and 200. The random test data proportion was 0.2 for the internal validation. The default settings were used for stopping conditions including minimum number of cases, maximum number of levels, minimum number in child node and the maximum number of nodes of 6, 10, 5 and 100, respectively. The best model was selected based on the estimation error for the internal test data.

4.3. Results of QSAR Models for Biliary Excretion

A total of 387 2D (e.g. kappa shape indexes, molecular connectivity indexes and electrotopological state indexes) and 3D molecular descriptors were used for the QSAR model development. Out of 217 compounds in the rat biliary excretion dataset, 9 compounds had excretion rate of 0%, and hence log BE% could not be calculated for them. The method of data allocation into training and test sets outlined above ensured that a similar biliary excretion and molecular property spaces were covered by both the training and the validation sets. BE% values ranged between 0.048 and 100 with mean log BE% values for the training and validation sets at 1.04 and 1.01, respectively. LogP was between -3.44 and 18.8 for the training set, and -3.17 and 7.83 for the validation set with similar mean values of 1.81 and 1.83, respectively. Molecular weights of the compounds were between 122 and 1215 Da for the training set and 94 and 1255 Da for the validation set, with mean values of 457 and 390, respectively. Scores plot from principle component analysis using all the molecular descriptors also indicates similar chemistry space for the two sets (Figure 4.1).

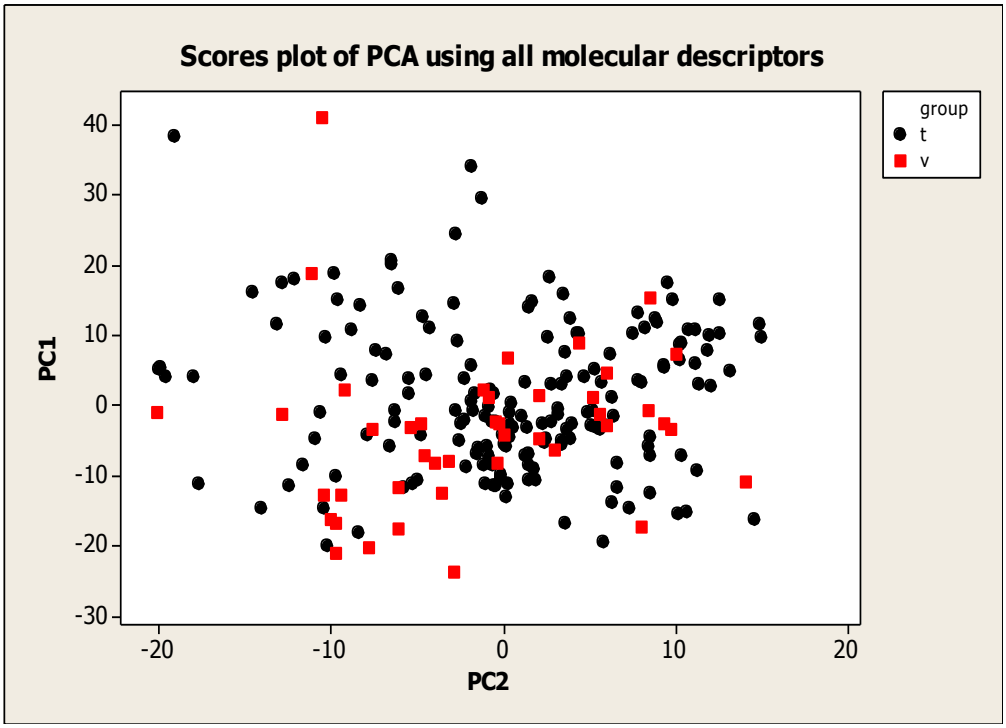

Figure 4.1. Scores plot of PCA using all 387 molecular descriptors

4.3.1. Regression Models

Linear regression equations are the simplest and most straightforward QSAR techniques. This has the benefit of easy interpretation which can provide some mechanistic insight into the process under investigation (Patel *et al.*, 2002). Stepwise regression analysis using *in vivo* rat biliary excretion data as the dependant variable resulted in the MLR (1) model below in which the number of molecular descriptors is limited to eight. The statistical terms of the equation are N the number of compounds, R-Sq the correlation coefficient, S the standard deviation and F Fisher's statistics and the P value. Observed versus calculated log BE% by this equation has been plotted (Figure 4.2.), with training and validation sets identified in the plot.

MLR (1) model

Log BE% = 2.09 + 0.00129 Vsurf_HB4 - 9.33 PEOE_RPC+ - 0.0574 SsCH3 - 0.377 fU - 0.00503 SlogP_VSA0 - 0.0573 SsssCH + 0.0403 AM1_dipole + 0.378 SddssS_acnt

N = 168 S = 0.489 R-Sq = 0.608 F = 30.9 P = 0.000

Molecular descriptors of this equation are not intercorrelated ($R^2 < 0.4$).

Table 4.2 gives a brief description of molecular descriptors used in this model. Vsurf_HB4 is the first molecular descriptor selected by the analysis and it indicates that compounds with high H-bond donor capacity have higher biliary excretion level. AM1_dipole (dipole moment) is the other polarity descriptor which has a positive effect. On the other hand, the equation shows that drugs with greater relative positive partial charge (PEOE_RPC+) have lower biliary excretion. The value of this descriptor is large for small acidic molecules such as benzoic acid and salicylic acid, and therefore the small size of such compounds may be the reason for the reduced biliary excretion. In this equation, fU with a negative coefficient indicates that compounds with higher unionised fraction at pH 7.4 have lower biliary excretion. In other words, although according to fU, acidity and basicity (dissociation in general) increase the biliary excretion of compounds, this is true only for large dissociated molecules. The positive effect of polarity and dissociation on biliary excretion is in agreement with the literature, where for example polar surface area (Gandhi and Morris, 2012) and an acidity indicator (Luo *et al.*, 2010; Chen *et al.*, 2010) have been included in linear QSAR models.

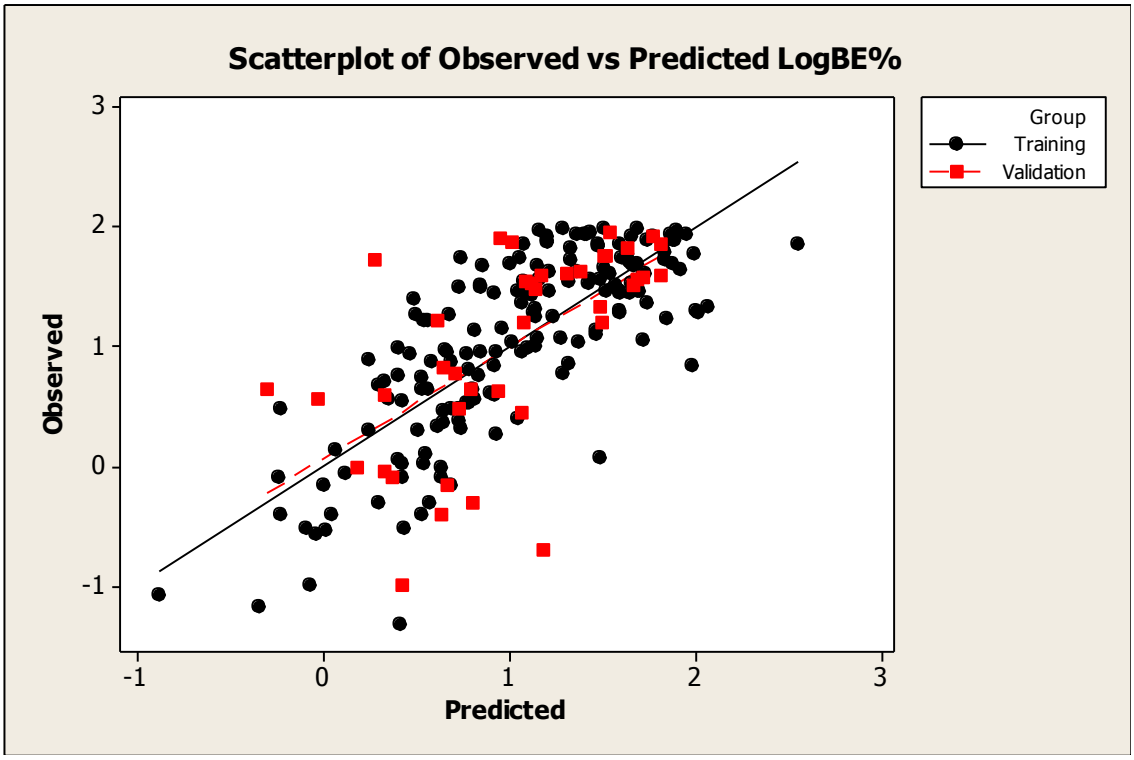

Figure 4.2. Observed vs predicted log BE% using MLR (1)

Also according to this equation, compounds containing many methyl groups (SsCH3) and those that are highly branched containing >CH− groups (SsssCH) have lower biliary excretion. Examples are macrolid antibiotics (i.e. telithromycin, azithromycin, erythromycin, actinomycin) muscle relaxant pipecuronium and the chemosensitizer PSC 833. The predominant excretion routes in these compounds are metabolism (Lee and Lee, 2007; Amacher *et al.*, 1991; Lam *et al.*, 2006; Lahiri *et al.*, 1970; Vereczkey and Szporny, 1980; Song *et al.*, 1998) except for pipecuronium and azithromycin for which the main excretion route is renal and biliary excretion respectively.

In this equation, SlogP_VSA0 shows the negative impact of the presence of atoms with LogP(o/w) contribution of less than or equal to -0.4. SddssS_acnt indicates the direct effect of sulphate or sulphonamide groups. Sulphate and sulphonamide groups are found in sulphonamide drugs such as succinylsulphathiazole, dyes such as methyl orange and sulphate conjugates such as estrone 3-sulphate which may be substrates of MRP2 or BCRP (Zamek-Gliszczynski *et al.*, 2006).

Table 4.2. A brief description of the most important molecular descriptors selected and used by the models.

Descriptor	Model	Description
a_acc	RF (1)	Number of H-bond acceptor atoms.
a_hyd	BT (1)	Number of hydrophobic atoms.
AM1_dipole	MLR (1), RT (1)	Dipole moment calculated using AM1 Hamiltonian.
BCUT_PEOE_0	RF (1)	The BCUT descriptor calculated from the eigenvalues of a modified adjacency matrix. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported, in this case the 2/3-ile. The diagonal takes the value of the PEOE partial charges.
CASA-	I-tree (2)	Negative charge weighted surface area, ASA- times max $\{q_i < 0\}$.
chi1	RF (1)	First order molecular connectivity index (Hall *et al.*, 2007).
COOH	I-tree (2)	Indicator variable for the presence of carboxylic acid group in the molecular structure.
Docking energy (MOE)	RF (1)	Docking score (kcal/mol)for enzyme-ligand docking of the compounds into the active site of P-glycoprotein (Aller *et al.*, 2009) calculated using MOE software
FASA_H	RT (1)	Fractional ASA_H calculated (water accessible surface area of all hydrophobic atoms) as ASA_H / ASA.
FCASA-	I-tree (1)	Fractional CASA- calculated as CASA- / ASA.
fU	MLR (1), RT (1)	Fractions of compounds unionised.
GCUT_SLOGP_1	RT (1)	The GCUT descriptors are calculated from the eigenvalues of a modified graph distance adjacency matrix. Each ij entry of the adjacency matrix takes the value 1/sqr(d_{ij}) where d_{ij} is the (modified) graph distance between atoms i and j. The resulting eigenvalues are sorted and the smallest, 1/3-ile, 2/3-ile and largest eigenvalues are reported. The diagonal takes the value of the atomic contribution to logP.
Kier2	BT (1), BT (2)	Second order kappa shape index: $(n-1)^2 / m^2$ (Hall *et al.*, 2007)
Kier3	BT (1), BT (2)	Third order kappa shape index: $(n-1)^2 / m^2$ (Hall *et al.*, 2007)
KierA1	I-tree (1)	First order alpha modified shape index: s $(s-1)^2 / m^2$ where s = n + a (Hall *et al.*, 2007)
KierA3	BT (1), BT (2)	Third order alpha modified shape index: (n-1) $(n-3)^2 / p3^2$ for odd n, and (n-3) $(n-2)^2 / p3^2$ for even n where s = n + a (Hall *et al.*, 2007).

Descriptor	Model	Description
LogD (5.5)	BT (2)	Logarithm of distribution coefficient D of a compound between octanol and buffer layers at pH value 5.5.
LogD (6.5)	RT (1), I-tree (1), BT (1), BT (2)	Logarithm of distribution coefficient D of a compound between octanol and buffer layers at pH value 6.5.
LogD (7.4)	BT (1), BT (2)	Logarithm of distribution coefficient D of a compound between octanol and buffer layers at pH value 7.4.
LogD (10)	I-tree (2), BT (2)	Logarithm of distribution coefficient D of a compound between octanol and buffer layers at pH value 10.
MW	I-tree (1) RF (1)	The molecular weight.
N ratio	RT (1)	The weight ratio of nitrogen atoms in the molecule.
PEOE_PC-	I-tree (2)	Total negative partial charge.
PEOE_RPC+	MLR(1), BT (2)	Relative positive partial charge: the largest positive atomic charge divided by the sum of the positive partial charges.
PEOE_VSA_NEG	I-tree (1)	Total negative van der Waals surface area.
PEOE_VSA-0	RT (1)	Van der Waals surface area of atoms with atomic charge in the range [-0.05, 0.00).
PEOE_VSA_FPPOS	RF (1)	Fractional positive polar van der Waals surface area. This is the sum of the VDW surface area such that partial charge of atom is greater than 0.2.
PEOE_VSA_HYD	BT (1), BT (2)	Total hydrophobic van der Waals surface area. This is the sum of the van der Waals surface area such that absolute value of atomic charge is less than or equal to 0.2.
Q_PC+	RF (1)	Total positive partial charge: the sum of the positive partial charge of atoms in the molecule.
SddssS_acnt	MLR (1)	Count of all sulphur atoms (ddssS) E-state values in molecule.
SlogP_VSA0	MLR (1)	Sum of approximate accessible van der Waals surface area for atoms with atomic contribution to logP(o/w) of equal or less than -0.4.
SMR_VSA7	I-tree (1)	Sum of approximate accessible van der Waals surface area for atoms with atomic contribution to molar refractivity of $R_i > 0.56$.
SsCH3	MLR (1)	Atom type electrotopological state index (sum of the E-states) for (-CH3) groups.
SsssCH	MLR (1)	Sum of E-State for all (>CH-) groups in molecule.
SsssC	I-tree (1)	Sum of all (> C <) E-State value in molecule.
TPSA	RF (1)	Topological polar surface area ($Å^2$).

Descriptor	Model	Description
VAdjEq	RF (1)	Vertex adjacency information (equality): This is an atom count /bond count descriptor calculated as: $-(1-f)\log_2(1-f) - f \log_2 f$ where $f = (n^2 - m) / n^2$, n is the number of heavy atoms and m is the number of heavy-heavy bonds. If f is not in the open interval (0,1), then 0 is returned.
vsa_hyd	BT (1)	Approximation to the sum of VDW surface areas of hydrophobic atoms ($Å^2$).
vsurf_CW4	I-tree (1)	Capacity factor is the ratio of the hydrophilic surface over the total molecular surface, calculated at eight different energy levels (from -0.2 to -6.0 kcal/mol).
vsurf_EDmin3	RT (1)	The lowest hydrophobic energy.
vsurf_HB4 vsurf_HB5 vsurf_HB6	MLR (1), BT (1), BT (1)	H-bond donor capacity at -2.0 Kcal/mol with carbonyl oxygen probe.
vsurf_ID7	RT (1)	Hydrophobic integy moment (The "integy moment" is defined in analogy to the dipole moment and describes the distance of the centre of mass to the barycenter of hydrophobic regions). Small integy moment indicates that the hydrophobic moieties are either close to the centre of mass or they balance at opposite ends of the molecule, so that their resulting barycentre is close to the centre of the molecule. VolSurf computes ID at eight different energy levels (from -0.2 to 1.6 Kcal/mol).
vsurf_IW2	I-tree (1), BT (2)	Hydrophilic integy moment (see vsurf_ID7).
vsurf_W1 vsurf_W3	RF (1), RT (1)	Hydrophilic volume.

4.3.2. Regression Tree Models Using C&RT

Several RTs were generated using a combination of molecular descriptors while cross-validation was applied. The best RTs were selected based on the standard error for the internal test set. As seen in Table 4.3, in RT (1), molecular descriptors were selected by C&RT analysis, while in I-tree (1), the molecular weight and in I-tree (2), the number of carboxylic acid groups were manually imposed as the first split descriptor using interactive C&RT routine in STATISTICA. These models

were developed using the training set while the validation set remained external. The RTs resulting from these trials have been presented in Figs. 4.3, 4.4 and 4.5. In the regression trees, N is the number of compounds, Mu is the average and Var is the variance of log BE% in each node. The molecular descriptors employed in the trees have been explained in Table 4.2.

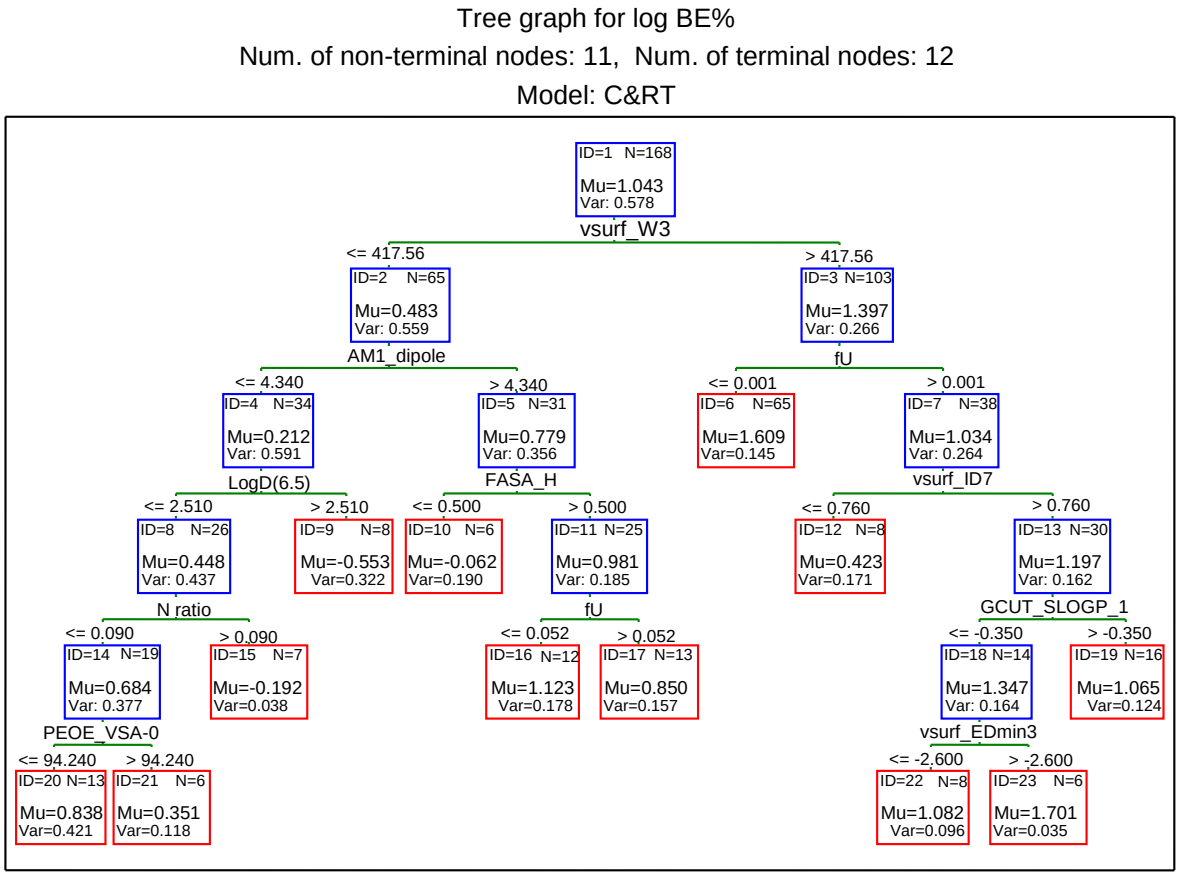

Figure 4.3. RT (1) developed using the training set with the descriptors selected by C&RT

Table 4.3 provides the statistical parameters of the regression trees.

Table 4.3. Description of the Regression Trees

Model No	Manually incorporated variables
RT (1)	None
I-tree (1)	Molecular weight
I-tree (2)	Carboxylic acid group

According to RT (1), biliary secretion is much higher for compounds with large hydrophilic volume (vsurf_W3), especially if they are ionised with $fU \leq 0.001$ (negligible unionised fractions at pH 7.4). Within the hydrophilic drugs of higher fU values (node 7), those with higher separation of lipophilic interaction sites from the centre of mass (vsurf_ID7 > 0.760) have higher biliary excretion. Surfactant molecules and glucuronide conjugates are examples of such molecules with high VolSurf integy moment (vsurf_ID7) and high biliary excretion. This branch follows to partition the molecules further according to GCUT_SLOGP_1 with compounds of lower hydrophobicity (node 18), and large hydrophobic interaction energy minima (vsurf_EDmin3 > -2.60) showing high biliary excretion (node 23). According to RT (1), the less hydrophilic drugs with vsurf_W3 values below 417.56 can be excreted heavily through the bile if they are highly dipolar (AM1-dipole > 4.336) with high ratio of lipophilic to total surface area (FASA_H > 0.50), especially if they are predominantly in the ionised form at pH 7.4 ($fU \leq 0.052$). On the other hand, compounds with low dipole moment have low biliary excretion specially if they are lipophilic with LogD(6.5) > 2.51 (node 9) or otherwise if they contain a high ratio of nitrogen atoms in the molecular structure (node 15). N ratio is low for larger alkaloids such as morphine or non-basic compounds, such as estrone 3-sulphate, which will have moderate biliary excretion especially if they are hydrophilic ($PEOE\_VSA\text{-}0 \leq 94.24$).

I-tree (1) was a result of molecular weight being employed in the first split using the interactive C&RT analysis in STATISTICA (Figure 4.4). The statistically selected molecular weight threshold was 347.9 Da, with the compounds below this weight showing lower log BE% values than the larger compounds. The tree shows that large ($MW > 347.9$) hydrophilic compounds (vsurf_CW4 > 0.540) have higher biliary excretion, particularly those with large total negative van der Waals surface area (PEOE_VSA_NEG) and low surface area corresponding to highly polarisable groups (SMR_VSA7), especially if they are highly branched (SssssC > -1.812). Within this group of compounds, larger molecules with KierA1 > 21.135 will have even higher biliary excretion. Other parameters of I-tree (1) indicate that high hydrophilic integy moment (vsurf_IW2) (node 13) and fractional negative charge weighted surface area (FCASA-) (node 11) would result in high log BE% value.

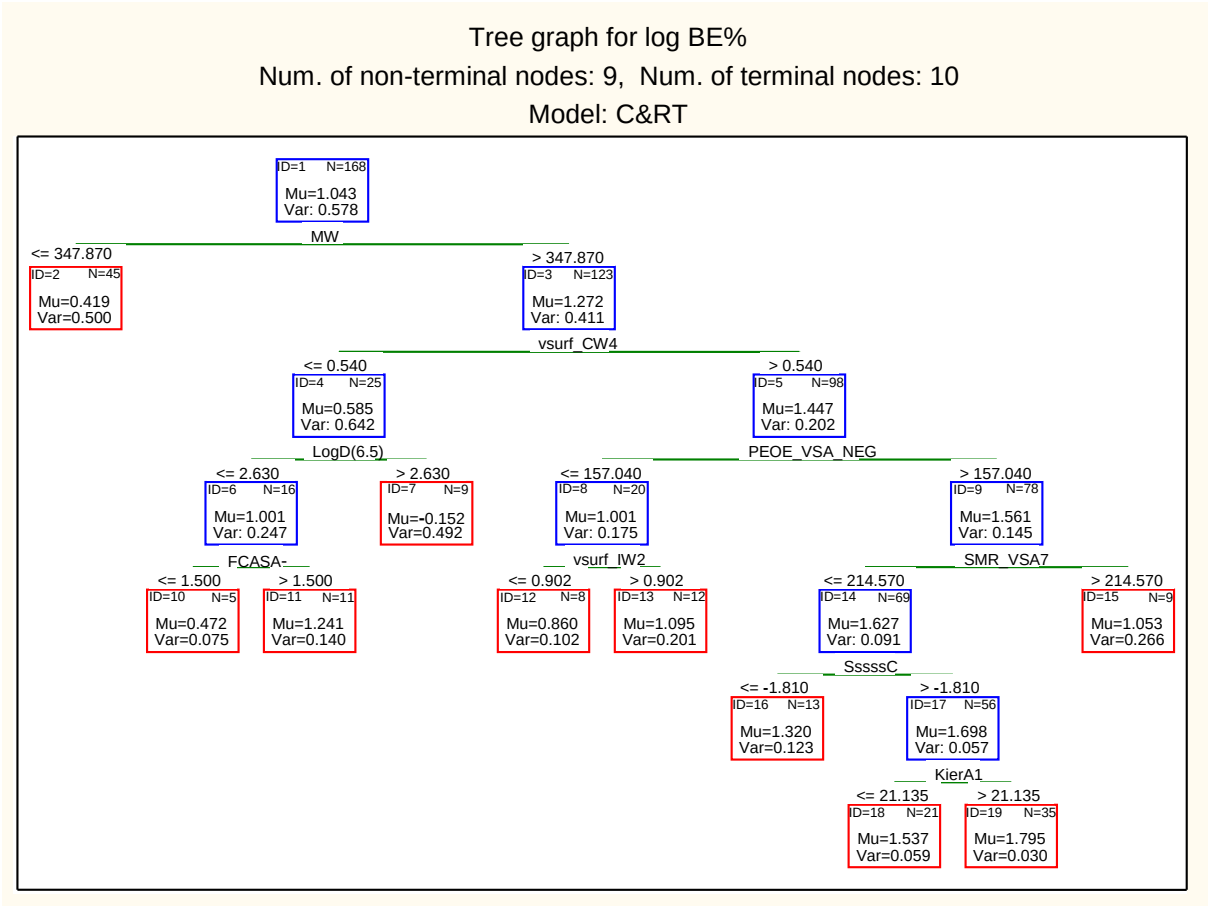

Figure 4.4. I-tree (1) developed using interactive C&RT analysis using molecular weight as the first descriptor.

Recent studies by Yang and co-workers show that presence of carboxylic acid group(s) may indicate a trend towards increased biliary excretion (Yang *et al.*, 2009). Therefore, the impact of presence of carboxylic acid group was examined using the interactive C&RT analysis with COOH used as the first partitioning molecular descriptor (Figure 4.5). According to I-tree (2), compounds containing at least one carboxylic acid group have higher biliary excretion levels. Furthermore, I-tree (2) indicated that compounds with lower total negative partial charge (PEOE_PC-) have much higher biliary excretion (node 6). These are large hydrophilic compounds with many negatively charged atoms. Non-acidic compounds in node 2 will have high biliary excretion if the negative-charge weighted surface area for these molecules is high (node 5). CASA− has an element of size as well as indicating the presence of negatively charged groups such as sulphates.

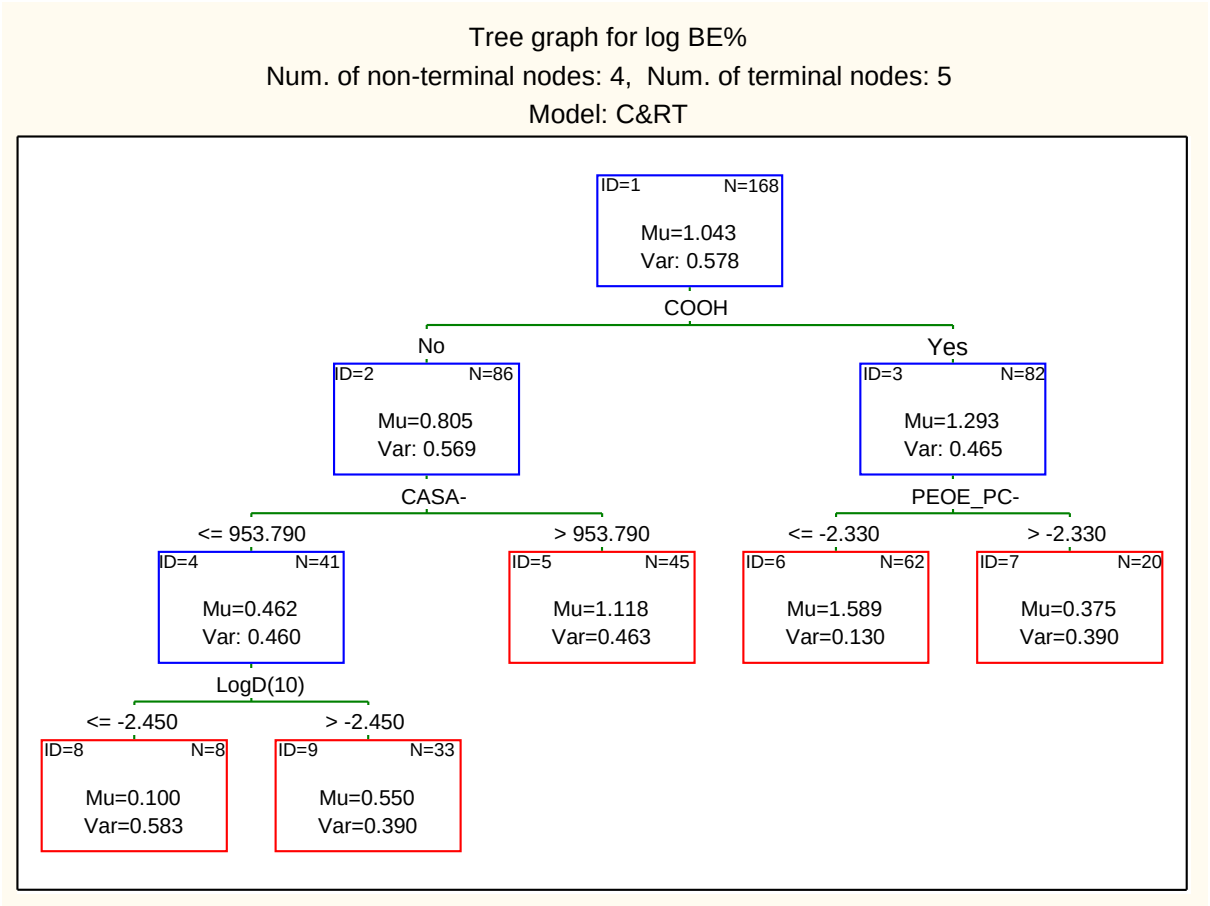

Figure 4.5. I-tree (2) using the number of carboxyl groups (COOH) as the first descriptor

Table 4.4. Statistical parameters of the models for training and test sets; RT is regression tree; BT is boosted trees and RF is random forest model

Model	Group	Risk Estimate	Standard Error
RT (1)	Train	0.112	0.040
	Validation	0.583	0.116
I-tree (1)	Train	0.229	0.034
	Validation	0.348	0.081
I-tree (2)	Train	0.323	0.050
	Validation	0.349	0.075
BT (1)	Train	0.079	0.007
	Validation	0.328	0.103
BT (2)	Train	0.078	0.007
	Validation	0.329	0.107
RF (1)	Train	0.262	0.047
	Validation	0.311	0.076

4.3.3. Boosted Trees

Boosted tree module computes a sequence of simple trees, where each successive tree is built for the prediction of the residuals of the preceding trees. The analysis using various combination of model parameters resulted in two best models selected based on the error level for the internal test set (Table 4.3). In models BT (1) and BT (2), the optimal numbers of trees were 145 and 147, with the learning rate of 0.10 and subsample proportions of 0.55 and 0.60, respectively.

It is possible to elucidate the influential descriptors in boosted trees analysis using variable importance calculation. Variable importance in STATISTICA is calculated as the relative (scaled) average value of the predictor statistic over all trees and nodes; hence these values reflect on the strength of the relationship between the predictors and the dependent variable of interest, over the successive boosting steps (STATISTICA help file, 2009). Included in Table 4.2 are the top ten most important molecular descriptors of BT (1) and BT (2) models. Some of the descriptors used by BT models are those already observed in RT and MLR models. For example, LogD (6.5) is present in two RT models and it is amongst the top ten most significant descriptors of both BT models. Other descriptors selected by these models are topological/size descriptors (KierA3, Kier2 and Kier3) and other lipophilicity descriptors such as LogD at different pH values and vsurf descriptors. Table 4.4 shows the statistical significance of these models. Graphs of average squared error against number of trees for training and cross-validated test sets can be found in Figures 4.6 and 4.7 for BT (1) and BT (2).

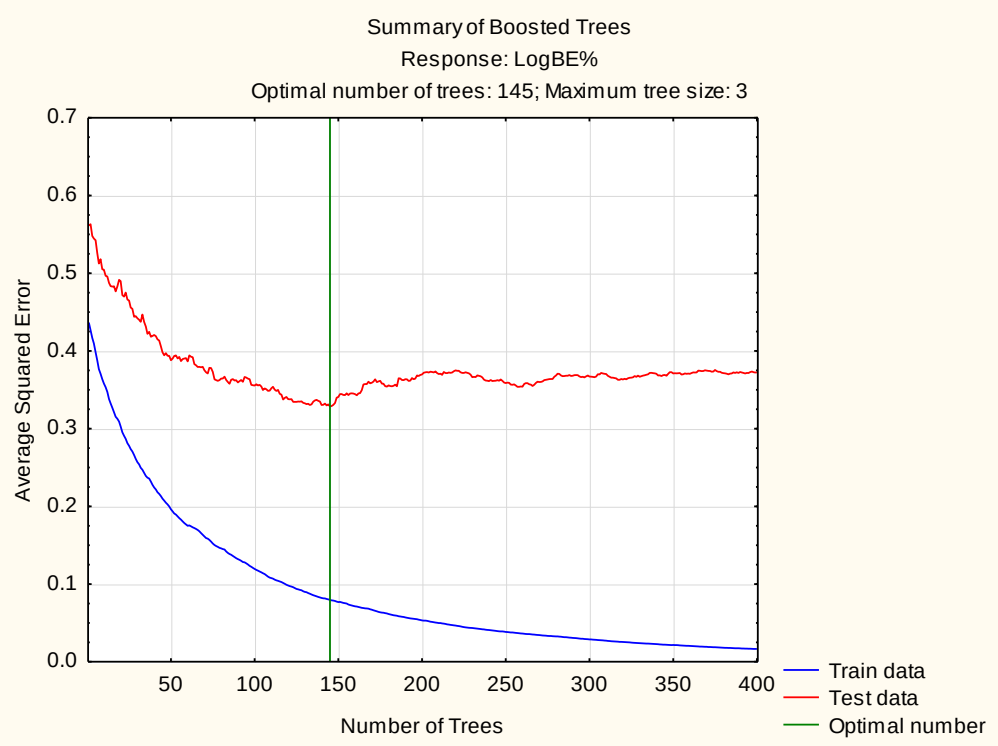

Figure 4.6. Average squared error of log BE% against the number of trees in the boosted trees model BT (1) for the training and internal test set

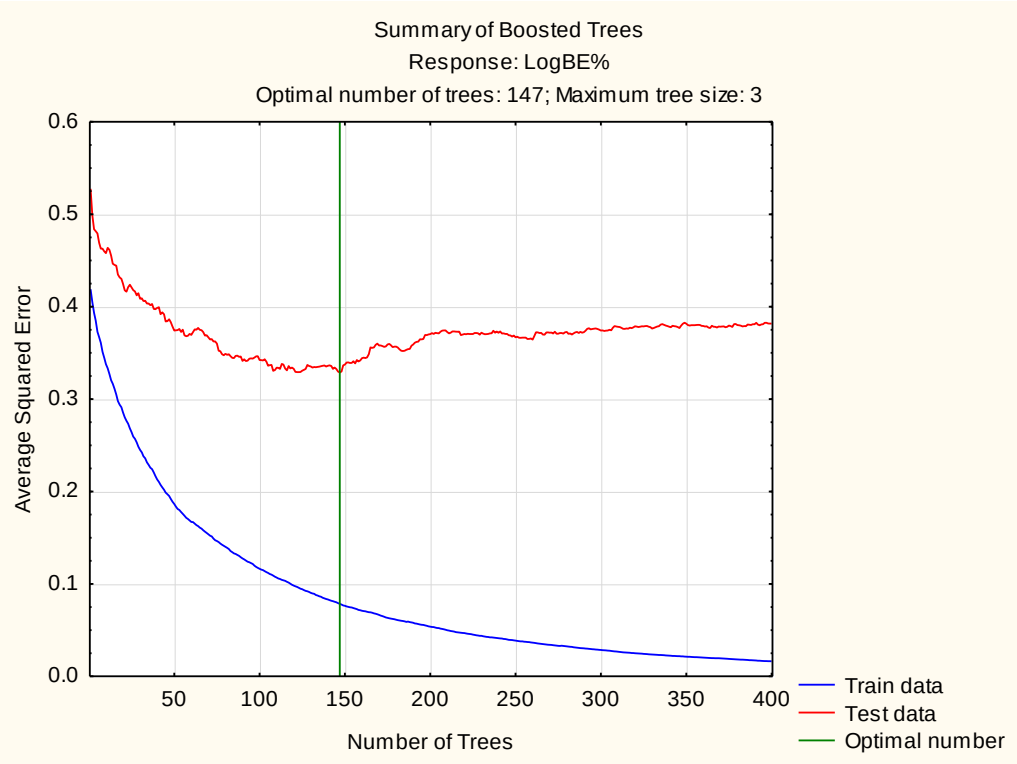

Figure 4.7. Average squared error of log BE% against the number of trees in the boosted trees model BT (2) for the training and internal test set

4.3.4. Random Forest

In RF, the number of trees specifies the number of simple regression trees to be computed in successive forest building steps. The model development used the default values of the software with the number of trees set at 100. The graph of average squared error against number of trees for training and cross-validated test sets indicates that the test and training set errors reach a plateau at around 10–15 trees (see Figure 4.8). The best model was achieved with a subsample proportion of 0.60, random test data proportion of 0.2 and number of trees of 100.

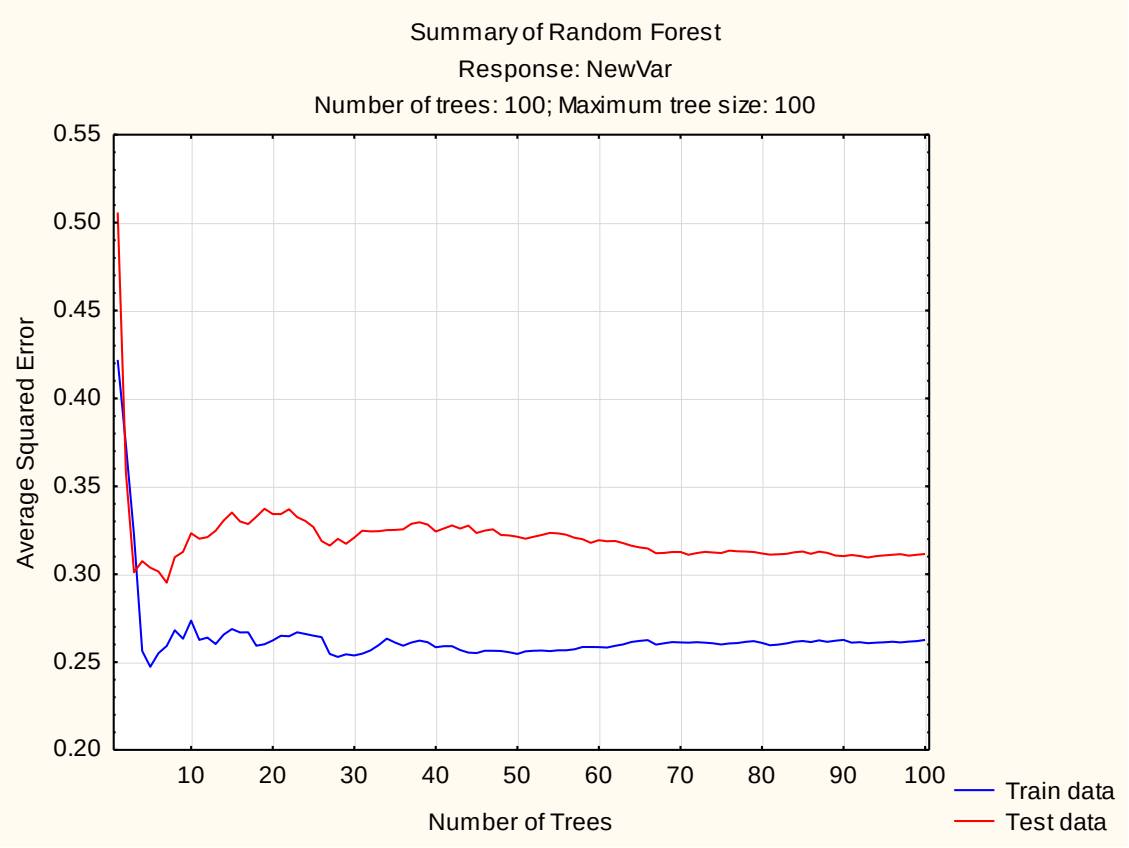

Figure 4.8. Average squared error of log BE% against the number of trees in the random forest model (RF) for the training and internal test set

Table 4.2 includes a description of the ten most significant descriptors employed in this model. Table 4.4 gives a summary of the statistical parameters of the RF model.

4.3.5. Validation of the Models

All models were validated by the same external validation set which had been set aside and not used at any stage of model development. Table 4.5 shows the prediction accuracy of the QSAR models using external validation in terms of the mean absolute error and the number of outliers. In addition, an average estimate of log BE% using all regression trees (RT (1) – I-tree (2)) was calculated and compared with the observed values to investigate any possible improvements in prediction accuracy. Table 4.5 gives the performance of this estimation method (consensus RTs).

Table 4.5. Summary of the prediction accuracy of the QSAR models

Model	MAE for training set	MAE for validation set
MLR (1)	0.377	0.483
RT (1)	0.304	0.373
I-tree (1)	0.345	0.451
I-tree (2)	0.424	0.468
Consensus RTs	0.319	0.383
BT (1)	0.229	0.412
BT (2)	0.226	0.417
RF (1)	0.403	0.496

4.4. Discussion

Biliary excretion can play a significant role in the elimination of drugs, and, therefore, its prediction is an important target in drug discovery. In the pharmaceutical industry, drug candidates are routinely tested in animal studies to measure the extent of biliary excretion and propensity of enterohepatic cycling, which have significant roles in the pharmacokinetics of a drug. In drug discovery, a reliable, user friendly and low-cost model based on computer-generated molecular

properties can reduce the number of high-cost animal (mainly rat) studies. This investigation aimed to elucidate how secretion into bile of compounds is controlled by their molecular structure, and to develop predictive models based on the molecular structure. Linear regression analysis, regression trees and two ensemble methods, boosted trees and random forest, were used for the QSAR model development.

4.4.1. Comparison of the Models

Linear regression equation is one of the simplest and the most common QSAR techniques. This method has the benefit of easy interpretation and it can provide mechanistic insight into the process under investigation. However, it has been argued that many biological processes have more complex relationships with the molecular attributes of the compounds and hence linear regression models may fail to capture these (Guha and Jurs, 2004). RT offers a suitable alternative to MLR method with the advantage of being flexibly non-linear while retaining the interpretability (De'ath and Fabricius, 2000). Ensemble methods such as random forest (Breiman, 2001) provide consensus predictions which may have improved accuracy. But this is often accompanied by a loss of interpretability, as the ensemble of many models is often used as a 'black box' prediction tool. In this investigation, STATISTICA variable importance analysis was used to find the most significant molecular descriptors in the boosted trees and random forest models.

According to Table 4.5, the most predictive model with the lowest estimation error for the external validation set is RT (1) followed by BT (1) and BT (2) and then I-tree (1). In other words, increasing the complexity of the models by allowing non-linear relationships and an ensemble of such models has been able to improve the prediction accuracy in comparison with a simple linear regression model (MLR). Table 4.5 shows the number of outliers from each of the models. According to this table, RT (1) followed by BT (2) and BT (1) and then I-tree (1) are the best externally validated models with the lowest numbers of outliers in the validation set. The advantage of RT is the obvious simplicity and interpretability which can make it more popular with the end users in drug discovery disciplines. For example,

when using the tree for a new compound, the molecular descriptors used in the tree will need to be calculated for the compound and then the terminal node (leaf) where the compound falls according to the molecular descriptor values should be identified. The average log BE% of the terminal node (Mu) is the estimate of the tree for this compound. Despite that RT provides discrete predictions of a continuous observation which is not ideal, this is a much more straightforward procedure than using BT or RF for the estimation of BE%. These models are ensemble of many trees, and therefore the prediction has to be performed by the computer rather than manually.

An interesting observation was made as MW and COOH were not significant in MLR equation when forced into stepwise regression analysis ($P > 0.05$). Despite this, incorporation of these two parameters was statistically significant in C&RT analysis resulting in I-tree (1) and I-tree (2). This indicates the non-linear nature of the impact of these two parameters on biliary excretion. Average prediction by the three RT models was also considered and found to be of similar accuracy to RT (1) (Table 4.5).

In this work, the MLR model based on the training set of 168 compounds had the second poorest prediction accuracy after RF. Studies by Yang *et al.* (Yang *et al.*, 2009) and Chen *et al.* (Chen *et al.*, 2010) report MLR models based on training sets of 37 and 46 compounds, respectively. The proposed model by Yang *et al.* incorporated molecular connectivity indexes and atom-type electrotopological indexes which have also been used in this study. The model proposed by Chen *et al.* also incorporated similar molecular descriptors to our study, with the addition of Abraham descriptors representing polarisability and hydrogen bond acceptor capacity. Although we have not used Abraham's descriptors, there are other molecular descriptors in our set of 386 descriptors that measure the same properties. Examples are the number of hydrogen bonding acceptor atoms and atomic charge on the most negatively charged atom in the molecule which may represent hydrogen bond acceptor ability (Dearden and Ghafourian, 1999) and molar refractivity descriptors which may indicate molecular polarisability (Verman and Hansch, 2005).

In another study, Luo *et al.* (Luo *et al.*, 2010) used 50 proprietary compounds from Bristol-Myers Squibb Co. for model development. They also developed a multiple linear regression model, but in addition to more common molecular descriptors, they employed free energy of aqueous solvation calculated from a self-consistent reaction field method. In analysing this model, Gandhi and Morris (Gandhi and Morris, 2012) found that the model failed to generalise further to the new set of compounds and specifically free energy of aqueous solvation was not statistically significant. They argued that a complex process such as hepatobiliary excretion cannot be captured by simple physicochemical properties when examining chemically dissimilar compounds. Indeed, such extrapolations to external compounds will fail when the compounds are outside the domain of applicability of the QSAR models. Incorporation of a larger dataset in this work may provide the opportunity for capturing an extended chemical space. This will be discussed further when analysing the outliers in the next two sections.

4.4.2. Structural Features of Compounds for Biliary Excretion

Table 4.2 gives a brief description of the significant molecular descriptors used in the models. For the sake of this discussion, the descriptors in this work can be classified roughly into five categories as follows: lipophilicity, ionisation, molecular size and topological and constitutional descriptors.

It can be seen in Table 4.2 that lipophilicity descriptors such as log D at different pH levels and surface area of hydrophilic molecules (SlogP_VSA0) are present in all models. In all interpretable models (except for the linear regression equation), lipophilicity descriptors show a negative effect on the biliary excretion of compounds. This may relate to the fact that highly lipophilic compounds are known to be highly extracted and metabolised in the liver (Proost *et al.*, 1997) rather than being excreted unchanged through bile or kidney. For example, metabolism by cytochrome P450 enzymes (Lewis and Ito, 2010) and (UDP)-glucuronosyltransferase (Smith *et al.*, 2003) is mainly controlled by lipophilicity and increased for more lipophilic compounds. There have been inconsistent findings in the literature regarding the effect of lipophilicity on the biliary excretion

of xenobiotics. Proost *et al.* found no significant correlation between lipophilicity and biliary excretion of a series of bulky organic cations despite it being the predominant factor for the degree of plasma protein binding and hepatic uptake rate (Proost *et al.*, 1997). Similar observations have been made for other compilations of biliary excretion data (Yang *et al.*, 2009). Other studies indicate negative effect of lipophilicity on the biliary excretion within the range of compounds studied (Luo *et al.*, 2010; Varma *et al.*, 2012). Lipophilicity has been associated with many models of ADME properties (Hansch *et al.*, 2004). It is a well-established fact that compounds with higher logP have poor aqueous solubility and are more likely to pass through lipid bilayer of biological membranes (Kerns and Di, 2008). The general trend in the literature with regards to the role of lipophilicity in pharmacokinetic processes indicates that more lipophilic compounds have higher oral absorption, plasma protein binding, and volume of distribution (van de Waterbeemd *et al.*, 2001; Obach *et al.*, 2008; Newby *et al.*, 2013b) and are more prone to P450 metabolism (Lewis and Ito, 2010; van de Waterbeemd *et al.*, 2001). This may lead to the reduced chance of excretion through bile as the intact drug.

All models presented in this work indicate the significant role of ionisation and polarity through molecular descriptors such as COOH, fU, FCASA− and SddssS_acnt. Acids are able to ionise into anions which are substrates of several transporters (generally organic anion transporters). Compounds that carry positive as well as negative charge or partial charges can use both the ‘organic anion’ and the ‘organic cation’ transport systems (Koepsell *et al.*, 2001). For example, OAT3 accepts various kinds of bulky hydrophobic anions, while OAT1 can transport relatively hydrophilic small molecules, such as nucleoside analogues (Maeda *et al.*, 2010). Besides, monocarboxylate transporters (MCT1 to MCT14) constitute a family of proton-linked plasma membrane transporters that carry molecules having one carboxylate group. MCT1 is expressed nearly all over in every tissue in the human body and also in rat and calves hepatocytes (Kirat *et al.*, 2007). MCT2 is abundant on the surface of human, rat and hamster hepatocytes (Halestrap and Meredith, 2004). MCT5 and MCT8 are also known to play transporting role in rat hepatocytes (Halestrap and Meredith, 2004). Studies of biliary excretion of exogenous compounds have indicated the relation between polarity and biliary excretion stating that possession of a strongly polar anionic group was important

factor in appreciable biliary excretion (Luo *et al.*, 2010; Millburn *et al.*, 1967). In all the interpretable models reported here, polarity descriptors show a positive impact on biliary excretion. Examples are the positive coefficients of dipole moment (AM1_dipole) in the linear regression equation and higher percent of compounds with lower unionised fractions at pH 7.4 (fU) in RT (1).

Molecular size is the other important factor in biliary excretion represented in the models by molecular descriptors such as kappa shape indexes, hydrophobic volumes (vsurf_W1 and vsurf_W3) and surface areas of atoms with specific charge or lipophilicity ranges (e.g. PEOE_VSA_NEG and PEOE_VSA_HYD). These molecular descriptors show positive effect on biliary excretion level in all models. This is in line with the common understanding that a molecular weight threshold may apply to biliary excretion of compounds, and that high molecular weight compounds may be predominantly excreted through bile (Yang *et al.*, 2009; Varma *et al.*, 2012; Millburn *et al.*, 1967). Yang *et al.* (Yang *et al.*, 2009) suggested a molecular weight threshold value of 400 Da for biliary excretion of anionic drugs in rats using 164 drugs. In this study, regression tree analysis found the threshold value for molecular weight to be at 347.9 Da for biliary excretion in rat (I-tree (1)). Incidentally, this regression tree had the second highest prediction accuracy for the external validation set amongst the RT models. This was despite the fact that molecular weight was not the descriptor of choice by C&RT analysis.

The incorporation of some structural fragments in the models gave interesting information regarding molecular requirements for biliary excretion. Examples include SddssS_acnt and SsssCH which indicate higher biliary excretion of compounds containing sulphate groups and non-branched structure (MLR). Compounds containing carboxylic acid groups are also more likely candidates for biliary excretion according to I-tree (2). Up to half of compounds in our dataset contain –COOH groups (103 compounds out of 217). Sixty-five out of 103 COOH containing compounds had biliary excretion of > 20%. Varma *et al.* (Varma *et al.*, 2012) have analysed the interconnection between physicochemical requirements of OATP substrates and the biliary excretion rates. It was then suggested that substrate specificity of OATPs including acidity may primarily indicate the elimination through bile (Varma *et al.*, 2012).

4.4.3. Analysis of the Outliers

There are a number of compounds that are outliers from majority of the models. Analysis of outliers may provide interesting information regarding the applicability of the models. Within the BE% range, it could be observed that compounds with low biliary excretion show a higher average error in general (Table 4.6). For example, the average error by all seven models was the highest for the six compounds with the extremely low biliary excretion (BE% < 0.23), followed by the compounds with 0.23 < BE% < 1.23 (-0.64 < log BE% < 0.09). A closer inspection of the data reveals that despite the high average error for the six compounds with low biliary excretion, the estimation may still be acceptable as all these compounds have been estimated to have a BE% value < 4% (average of all models) and below 0.6% by RT1 model with only one exception (benzoic acid). A hypothesis here could be that these compounds may have suitable properties for higher biliary excretion, but other routes of elimination are predominating. For example, it has been shown for benzoic acid that when clearance by the kidney is prevented, biliary excretion increases by 10% (Abou-El-Makarem *et al.*, 1967). Out of 217 compounds in the dataset, the predominant routes of elimination are biliary excretion for 115 compounds, renal excretion for 65 compounds and metabolism for 37 compounds. However, the outlier compounds do not belong to any single groups above in terms of the predominant routes of elimination (see Figure 4.9 for a graph showing the predominant routes of elimination for the compounds in biliary excretion dataset).

Table 4.6. Average MAE by nine models for compounds with various BE%, logP and molecular weight values

BE%	Average MAE	n
<= 0.23	1.12	6
0.23 - 1.23	0.50	26
> 1.23	0.30 [a]	176
MW (Da)		
> 280	0.31	173
<= 280	0.54	35

Log P		
> 5.35	0.63	13
<= 5.35	0.33	195

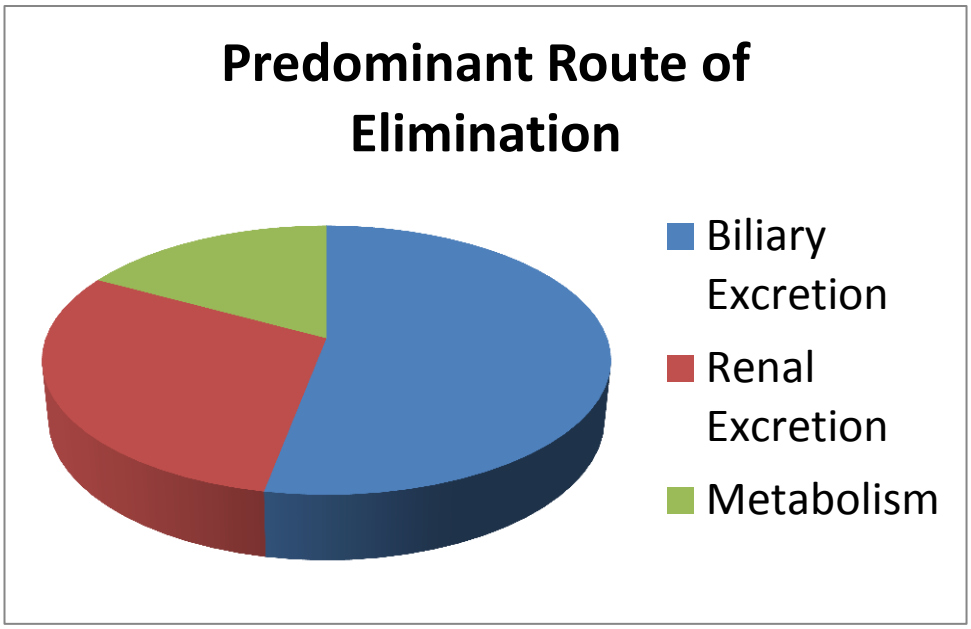

Figure 4.9. The main routes of elimination for compounds in the biliary excretion dataset

According to Table 4.6, highly lipophilic compounds (log $P > 5.35$) and low molecular weight compounds ($MW \leq 280$) also show higher error rates, and this may need to be considered when using the models for the prediction of external compounds.

Table 4.7 gives a list of the compounds that are outliers in six or seven models out of the seven models proposed here. In addition, there are four compounds which were outliers in four or five models but had exceptionally high average error from the seven models. These compounds were part of the training or validation sets, but none were omitted from average error calculations.

Table 4.7. Outlier compounds in training or validation sets with absolute error of > 0.6 in more than five out of seven models and their BE% values.

Outliers	BE%	Log BE%	Over or under prediction	Models with error	MW
Benzoic acid	0.09	-1.07	over-predicted except for BT	4	122
EMDP	0.20	-0.69	over-predicted	6	263
Fosmidomycin	0.10	-1.00	over-predicted	7	183
Nelfinavir	0.05	-1.32	over-predicted	5	567
EDDP	36.31	1.56	under-predicted	6	277
PAEB	31.62	1.50	under-predicted	7	222
Tolrestat	53.70	1.73	under-predicted	6	357

The outliers in Table 4.7 have been over- or under-predicted by the models. One compound in the table has shown underestimation by some and overestimation by other models; biliary excretion of benzoic acid was overestimated by all models except for BT (1) and BT (2). It can be seen in Table 4.7 that fosmidomycin, nelfinavir and 2-ethyl-5-methyl-3,3-diphenyl-1-pyrroline (EMDP) are over-predicted by five or more models. Benzoic acid is rapidly cleared by the kidney, so it may not have enough time to pass into the bile (Abou-El-Makarem *et al.*, 1967). Abou-El-Makarem and his colleagues examined this possibility by tying up the renal pedicles in rats, so that clearance by the kidney was prevented, and the results indicated that when clearance by the kidney is prevented, biliary excretion increased by 10% (Abou-El-Makarem *et al.*, 1967). Fosmidomycin has a short half-life of 1.7 h and is rapidly cleared by the kidneys (Murakawa *et al.*, 1982). It is a small molecular weight polar agent which may not be cleared in high quantities through bile according to the molecular weight threshold hypothesis. Despite the use of molecular size descriptors, this compound still appeared to be overestimated by all seven models, even using I-tree (1) which has employed MW for the first branching. The problem with I-tree (1) in relation to this compound is that although this compound falls into node 2 along with 44 other low molecular weight compounds, this node has an average log BE% of 0.42 which is much lower than node 3 with an average log BE% of 1.27 but not low enough for this compound. Likewise, other models have indicated low biliary excretion of small-sized compounds, but somehow, estimation is higher than what is actually observed.

Nelfinavir has a half-life of 3.5 to 5 h and is eliminated via metabolism by the cytochrome P450 enzyme system (Bardsleey-Elliot and Plosker, 2000). This is a highly lipophilic compound which is poorly excreted through bile, and is predicted as such by the models (predicted BE% below 2% using all models except for I-tree (2) and RF which predict 13 and 7.6%, respectively).

EMDP is a major metabolite of methadone which has been over-predicted by most models despite a very low biliary excretion. As with nelfinavir, the predicted BE% for this compound by most models is quite low at < 4% (MLR is an exception) and the selected model, RT (1), predicts a biliary excretion value of ~0.3%. Despite this, in comparison with the extremely low observed value of 0.05%, the predicted values are much higher, leading to a numerically large average error, even though qualitatively, the predicted biliary excretion may be reasonably low.

EDDP, PAEB (procaine amide ethobromide) and tolrestat are the under-predicted compounds. All these compounds have high BE% values at 36, 32 and 54%. This is despite the relatively low molecular weights of EDDP and PAEB which are below the defined MW threshold of 347 Da for biliary excretion. The exact mechanism of high biliary excretion of these compounds warrants further investigation to explore the reasons behind such high biliary excretion despite the low molecular weight.

Tolrestat has a relatively high molecular weight suitable for biliary excretion and a COOH group making it a suitable substrate for OATPs (Varma *et al.*, 2012). Despite this, the hydrophilic volume calculated by the VolSurf descriptor vsurf-W3 is not high enough to put this compound in node 3 rather than node 2 of RT (1) model. In I-tree (1), the compound falls into node 16, which is due to the lack of non-aromatic branched structure which would place it in node 17 with a higher predicted BE%. Likewise, in I-tree (2), this compound fails to be placed in node 7 and falls in node 6 instead due to the low total negative charge (> -2.33) as a result of the low number of negatively charged atoms. This indicates a shortcoming in the abovementioned models which lack suitable parameters that can capture the relative polarity in relation to the molecular size.

4.5. Conclusion

This investigation focused on the development of computational models for a cost-effective estimation of biliary excretion of compounds. This was made possible through the application of quantitative structure-activity relationships where molecular properties (descriptors) of a large dataset of compounds were related to the percentage of dose excreted intact via the bile through the use of statistical techniques. Some of the statistical techniques led to very promising results as evaluated by the prediction accuracy for the external validation set. The QSAR models also identified the important molecular properties (descriptors) that have the main influence on biliary excretion of compounds. The selected models were the regression tree (C&RT) model, RT (1), followed by boosted trees models BT (1) and BT (2). Regression trees also have the advantage of being simple, interpretable and user-friendly. The models generally indicated that larger, relatively hydrophilic molecules containing a carboxylic acid group are more prone to biliary excretion. For example, in the selected model, RT (1), compounds with increased hydrophilic volume and acidic dissociation have high biliary excretion. The significance of acidity and molecular size were further confirmed through interactive regression trees and a statistically validated MW threshold for effective biliary excretion was established. Detailed analysis of the error levels and outliers indicated that the models work best for larger compounds (MW > 280 Da) and are less accurate for extremely lipophilic compounds (log P > 5.35).

5. Effect of P-gp Binding on Biliary Excretion

5.1. Introduction

One in four deaths in the United States is due to cancer and recently the American Cancer Society reported a total of 1,660,290 new cancer cases and 580,350 cancer deaths are projected to occur in the United States in 2013 (Siegel *et al.*, 2013). The failure of cancer treatment can be attributed to a variety of different pharmacological and clinical reasons; but one major cause of the treatment failure is multidrug resistance (MDR) to chemotherapeutics (Song *et al.*, 2010). MDR mechanisms can result in resistance to a number of structurally and functionally unrelated chemotherapeutic agents. The multidrug resistance behaviour is mainly linked to the activity of transmembrane efflux pumps such as P-glycoprotein 1 (P-gp/ABCB1), breast cancer resistance protein (BCRP/ABCG2) and multidrug resistance-associated protein 1 (MRP1/ABCC1), which are members of ATP-Binding Cassette transporter family (Krishna and Mayer, 2000). P-gp, also known as multidrug resistance protein 1 (MDR1), is a well-studied glycoprotein which was first discovered in 1976 by surface labelling studies in multidrug resistant Chinese hamster ovary cells (Juliano and Ling, 1976). Since then, it has demonstrated its function as a transporter of hydrophobic drugs, lipids, steroids and metabolic products.

Overexpression of P-gp in cancer cells contributes significantly to the resistance of cancer cells against chemotherapeutic agents (Gottesman, 2002). As a strong efflux pump, P-gp is able to export a number of structurally diverse anticancer agents including anthracyclines, epipodophyllotoxins and vinca alkyloids. As a result, P-gp has been suggested as a viable target to be inhibited in the treatment of multidrug resistant cancer (Szakács *et al.*, 2006). Drugs such as actinomycin-D and azithromycin can strongly block the P-gp and limit the efflux of P-gp substrates. Inhibitors that block the transport of chemotherapeutics or other compounds may act as competitive or non-competitive inhibitors (Ambudkar *et al.*, 1999). In recent years, the inhibitory activity against P-gp has been tested in many compounds in

order to overcome P-gp mediated resistance of cancer cells to the chemotherapeutics (Pajeva *et al.*, 2009).

In addition to its role in multidrug resistance, P-gp has a profound role in pharmacokinetics, affecting drug absorption, distribution and excretion (Lin and Yamazaki, 2003). It is found in high amounts at the apical surface of epithelial cells lining the colon and small intestine, hepatocytes, pancreas ductules, proximal tubules in kidneys, and the adrenal gland (Schinkel and Jonker, 2003; Dean, 2002). P-gp is also known to play a major role in transporting compounds out of the brain in the blood brain barrier (Malmo *et al.*, 2013). In the BBB, only suitably lipophilic compounds can diffuse across the endothelial cells and enter the brain. However, a high proportion of P-gp that surrounds this area of the brain prevents their accumulation by distributing substrates back into the blood circulation (Malmo *et al.*, 2013). In the gastrointestinal tract and in hepatocytes, P-gp is responsible for the efflux of drugs back into lumen/bile, thus reducing the bioavailability of substrate drugs (Giacomini *et al.*, 2010). Similarly, in kidneys, P-gp is located primarily in glomerular mesangium and the apical membrane of proximal tubule epithelia and plays a significant role in the tubular secretion of organic cations (Giacomini *et al.*, 2010).

As stated earlier, P-gp is poly-specific and can efflux a very broad range of substrates. The substrates can have molecular weights ranging from 250 to 1850 Da, different ionization states, acid/base properties, hydrophobicities or amphipathic properties (Kerns and Di, 2008). There are drugs and herbal products that can affect the function of P-gp transporters and the number of drugs that are found to be the P-gp substrates is incessantly growing. For instance, rifampin (an antituberculosis drug) induces the intestinal expression of P-gp (Ehrhardt and Kim, 2008). Due to the broad substrate specificity of P-gp, drug-drug interactions involving P-gp are very likely (Lin, 2003). Drug–drug interaction is an important issue observed in cancer patients, especially because they often receive multiple medications concurrently with complex chemotherapy regimens (Wong *et al.*, 2008). Due to the importance of P-gp in drug interaction, the FDA has urged that every new molecular entity should be routinely checked for a possible interaction with P-glycoproteins (FDA Guidelines, 2014).

Multiple binding sites are available for P-gp. Generally, P-gp inhibition can happen in three different ways. Firstly by blocking the binding of substrate drugs this can be allosteric, competitive or non-competitive. Secondly by acting with ATP hydrolysis site, due to the fact that P-gp is inactive when ATP hydrolysis site is blocked (Shapiro and Ling, 1997; Urbatsch *et al.*, 1995). Although majority of drugs block the P-gp by blocking the substrate binding sites (Varma *et al.*, 2003), presence of multiple binding sites should be considered in the substrate or inhibitor studies. Besides, P-gp may be induced by various agents such as ritonavir (Perloff *et al.*, 2001).

Numerous well-known multispecific drug transporters are involved in liver canalicular efflux of many xenobiotics (Pfeifer *et al.*, 2014). Of these transporters, P-gp characterizes as the most widely studied efflux transporter in biliary excretion. This transporter is responsible for transporting of mainly large lipophilic and cationic substrates into the bile canalicular (Oza, 2002). It has been shown in genetically modified mice lacking mdr1-type (drug-transporting) P-gp that substrate drugs such as digoxin may have a reduced elimination (Schinkel *et al.*, 1997). Moreover, mutations in the human MDR3 gene responsible for P-gp lead to progressive familial intrahepatic cholestasis which lack biliary phospholipid excretion (de Vree *et al.*, 1998). Another example regarding the importance of P-gp in biliary excretion of drugs is the P-gp substrate imatinib, which shows a significantly reduced fecal excretion in P-gp knockout mice or in the presence of P-gp inhibitors (Oostendorp *et al.*, 2009).

Given the important clinical relevance of P-gp, it is important to elucidate the mode of interaction with the modulators and substrates of this enzyme. Higginis and colleagues suggested a model for the P-gp polyspecificity namely “hydrophobic vacuum cleaner” model (Higgins and Gottesman, 1992). In the proposed model, the hydrophobic substrates enter the transmembrane domain of P-gp and are transported outside the cell. A recent study by Aller et al (Aller *et al.*, 2009) provided a detailed structural description of mouse P-gp, which indicates a substantial internal cavity comprising mostly hydrophobic and aromatic residues. Despite the substrate promiscuity, several studies have been valuable in identifying structure activity relationships for the modulators. Evidences from X-ray

crystallography (Aller *et al.*, 2009), chromatography (Lu *et al.*, 2001) and several biochemical techniques (Martin *et al.*, 2000; Maki *et al.*, 2006) suggest the presence of multiple substrate-binding sites and a number of inhibition mechanisms, which may be the cause of substrate promiscuity. As a result, it may be necessary to generate more than one pharmacophore for P-gp (Ekins and Erickson, 2002).

The type of the quantitative data available for P-gp is mostly in terms of IC_{50} values for the inhibitors. On the other hand very few substrate Km measures are found in the literature, despite the availability of binary data of substrate/non-substrate (Matsson *et al.*, 2009). As a vast majority of the reported IC_{50} values are for compounds that also act as substrates, with the exception of flavonoids which are believed to be able to bind to the ATP site as well as the substrate binding site (Kim, 2002), the inhibition constants may also indicate the binding capacity of the compounds. As a result, in this investigation, the IC_{50} and K_i values were collated for the QSAR studies. The use IC_{50} (concentration of inhibitor required for 50% inhibition) has the disadvantage of not allowing easy comparison of data from different substrate conditions. Unlike IC_{50}, the inhibition constant, K_i, is a more universal parameter that is standardised according to the substrate concentration and Km values (Cheng and Prusoff, 1973). A K_i value is related to enzyme-inhibitor complex and explains the strength of the interaction.

The broad aim of this investigation was to study the effect of P-gp binding on the QSAR models for the estimation of P-gp. To achieve this, first, several data mining techniques were used to enable development of universal models for the prediction of P-gp inhibition constant (K_i). In these models, the use of molecular descriptors for the substrates in addition to the inhibitor parameters may be useful for splitting of the K_i data if the substrate type has an effect on the measured K_i values. Secondly, docking scores were investigated as a complementary parameter to investigate the significance of interaction energy between the ligands and P-gp in the models for estimation of the binding constants. Third, the selected QSAR models were used for the prediction of P-gp binding constants of the compounds in biliary excretion dataset. Finally, the predicted P-gp dissociation constant (briefly log K_i) values were used as predictors in the QSAR models for the prediction of biliary excretion.

5.2. Methods

5.2.1. P-gp Dataset

IC_{50} and K_i values for P-gp inhibitors were collated from the literature (Cook *et al.*, 2010; Choo *et al.*, 2000; Dantzing *et al.*, 1996; Eberl *et al.*, 2007; Ekins *et al.*, 2002a; Ekins *et al.*, 2002b; Eriksson *et al.*, 2006; Kakumoto *et al.*, 2002; Katoh *et al.*, 2001; Keogh *et al.*, 2006; Lan *et al.*, 1996; Lumen *et al.*, 2010; Luo *et al.*, 2002; Matsson *et al.*, 2009; Neuhoff *et al.*, 2000; Noguchi *et al.*, 2009; Pauli-Magnus *et al.*, 2000; Petri *et al.*, 2004; Rautio *et al.*, 2006; Richter *et al.*, 2009; Shaik *et al.*, 2007; Tang *et al.*, 2002a; Tang *et al.*, 2002b; Wandal *et al.*, 1999 and Wang *et al.*, 2001). IC_{50} values of P-gp inhibitors were used to calculate the K_i values using the Cheng-Prusoff equation below.

$$K_i = \frac{IC_{50}}{1+\frac{[S]}{K_m}} \qquad \textbf{Eq (1)}$$

In this equation [S] is the substrate concentration and K_m is Michaelis–Menten constant for the substrate (the concentration of substrate at which enzyme activity is at half maximal). If K_m values for the substrates were not reported in the publication, then they were obtained from the authors through personal communication. The rationale behind converting the IC_{50} values to K_i values is that the K_i is a more universal scale, which in theory should be independent of the substrate used.

In case there were several IC_{50}/K_i values available for a single inhibitor from different sources, the average K_i values were used, unless the probe substrate was different. If there was a significant difference in the reported IC_{50}/K_i values, we contacted the authors to find out if they could provide an explanation for the observed differences before using the reported values. In total the dataset consisted of K_i values for 219 unique inhibitor/substrate pairs, with data measured in different cell systems including Caco-2, MDCK-MDR1, MDCK II-MDR1, K562-MDR, MDR1 transfected LLC-PK1 and P388 lymphoma cells. Human colon carcinoma cell line (Caco-2) and Madin-Darby canine kidney cells (MDCK) were the most common cell line used in our dataset. The inhibitors in the dataset are from

different chemical/pharmacological classes such as anticancer and anti-HIV agents, statins, antiretrovirals, cephalosporines, ergopeptides, antipsychotics, opiods, NSAIDs, analgesics, and antiarithmetic drugs. The dataset is presented in Appendix II.

5.2.2. P-gp-Ligand Docking

Docking energy for all inhibitors was calculated using MOE software (version 2012.10, Chemical Computing Group Inc. Montreal, Canada). Later, the docking score of inhibitors were used as an additional molecular descriptor by adding these score's columns to the dataset.

The X-ray structure of the mouse P-gp was obtained from the protein data bank (PDB code 3G60) [http://www.rcsb.org]. The use of this PDB structure was due to a previous docking investigation that showed better scoring poses using mouse 3G60 structure in comparison with the other two mouse P-gp structures (PDB codes: 3G61 and 3G5U), or the human homology model of P-gp (Löschmann *et al.*, 2013). It should be noted that this structure of mouse P-gp was co-crystalised with a ligand and the complex had two stereo-isomers of cyclic hexapeptide inhibitors, cyclic-tris-(R)-valineselenazole (QZ59-RRR) and cyclic-tris-(S)-valineselenazole (QZ59-SSS) in the active site (Aller *et al.*, 2009). The protein was protonated and protonatable residues were titrated using default parameters of the software before the docking exercise. Molecular structures of the ligands (P-gp inhibitors) were optimised after atomic charge calculation using SCF optimization (AM1 Hamiltonian). In enzyme-ligand docking, default parameters of the software were used for ligand interactions. These are energy cut-off for H-bond and ionic interactions of -0.5 kcal/mol and maximum distance for non-bonded interactions of 4.5 Å. In the MOE dock panel, the placement method was Triangle Matcher, the scoring methodology was set to London dG as the first and the second scoring functions, the refinement methodology was set to Forcefield, and finally, the 30 best scoring poses, the mean energies and the mean energies and backbone root mean square deviation (RMSD) were retained. The binding site was defined in MOE software using the co-crystallised ligand QZ59-RRR.

Preparation of compounds for Docking

Before docking could take place, the SDF file was imported into the MOE software. MOE is a suite of applications that can be used to manipulate and analyse a collection of compounds. For docking to work efficiently, it is essential that each structure is in a form suitable for it to be docked to a ligand. As a result, the software's 'Wash' application was used to clean the structures and neutralise the protonation state of each compound. This will neutralise all atoms and form the structure of the compound in its least charge-bearing state. The next step was to lower the potential energy of the structures. This was completed using the "Energy minimize" function from the software. The compounds in the database were now ready to be computed and molecular descriptors were calculated.

Validation of docking experiment

The published X-ray crystallography structures (Aller *et al.*, 2009, Gutmann *et al.*, 2010) were used to validate our docking model by comparing the geometries of the docked Abcb1a/QZ59-RRR structure and the structure of the Abcb1a/QZ59-RRR complex from X-ray crystallography and measuring root-mean-square deviation (RMSD) between them.

5.2.3. Model Development and Validation

Development of models for P-gp

To perform QSAR analyses, P-gp inhibitors were divided into validation and training sets. To divide the inhibitors, they were ordered with ascending K_i values, and then from every five compounds, four were allocated into the training and one into the validation set randomly. This ensured similar K_i ranges for the validation and training sets. In this way, training data consisted of 176 compounds and external validation set consisted of 43 compounds.

In this study, QSARs were established to relate the P-gp binding effect of compounds (log K_i) to the molecular descriptors and P-gp docking scores. Molecular descriptors were calculated according to the procedures explained in

section 3.1. Before building the models, the molecular descriptors were checked to find and discard those columns containing more than 98% constant values or more than 10% missing values. The total number of molecular descriptors used in all statistical analyses was 388.

STATISTICA Data Miner version 11 was used for the statistical analysis. Statistical methods consisted of decision tree methods and ensemble methods including Classification and Regression Tree (C&RT), Chi-square Automatic Interaction Detector (CHAID), Boosted Trees (BT) and Random Forest (RF). Moreover, Multivariate Adaptive Regression Splines (MARS) model was also developed. These methods have been explained in Chapter 3. Log K_i was the dependent variable and the predictors were selected by the embedded feature selection methods in C&RT, CHAID, BT and RF from all the molecular descriptors and docking scores available for the inhibitors and substrates. In C&RT analysis, several stopping criteria were examined, including the default settings in STATISTICA. The default stopping criteria were minimum number of cases of 24 to allow further splitting, and the maximum number of nodes set to 100. The V-values of 10 or seven was used in the V-fold cross-validation. In CHAID analysis, STATISTICA default setting for stopping criteria were used, including minimum number of cases for splitting of 22, maximum number of nodes of 1000, probability for splitting of 0.05 and probability for merging of 0.05. In BT analysis, the default values for learning rate, the number of additive terms, random test data proportion and subsample proportion were 0.1, 200, 0.2 and 0.5 respectively. Various subsample proportions of 0.45, 0.50, 0.55 and 0.60 were also examined in combination with the learning rates of 0.10, 0.03, 0.05 and 0.08. In RF analysis, various subsample proportions of 0.45, 0.50, 0.55 and 0.60 were examined. The random test data proportion was 0.3 for the internal validation and number of trees was 100. The default settings were used for stopping conditions including minimum number of cases, maximum number of levels, minimum number in child node and the maximum number of nodes of 5, 10, 5 and 100, respectively.

For the development of MARS model, several pre-processing feature selection techniques were examined. Feature selection methods were a Chi-square method as implemented in STATISTICA v11 (StatSoft Ltd.) developed by Hill and Lewicki

(Hill and Lewicki, 2006), stepwise regression analysis, and variable importance rank from random forest and boosted trees analyses. The Chi-square-based feature selection in STATISTICA picks a subset of descriptors from the descriptor pool without assuming that the relationships between the predictors and the dependent variables are linear or even monotone. In this feature selection, the range of continuous variable values was divided into 10 intervals. The best variables picked by STATISTICA feature selection, the best descriptors selected by stepwise regression analysis, as well as the top 5, 10, 15, 20 and 25 descriptors picked by RF, and the top 5, 10 and 15 descriptors picked by BT were examined in separate MARS analyses and the resulting models were compared. In MARS analysis, the default model specifications for maximum number of basis functions, degree of interactions, penalty and threshold were 21, 1, 2 and 0.0005 respectively.

The best model from each analytical method was selected based on the performance indicators for the internal validation set.

Development of models for biliary excretion incorporating predicted P-gp activity

The selected P-gp dissociation constant (K_i) models above were used to predict the log K_i values for compounds in biliary excretion dataset (n = 217). QSAR models were developed for biliary excretion using the dataset and methods explained in Chapter 4. In addition to the molecular descriptors, the P-gp effects predicted by the selected models from section 5.2.3 were used as the independent variables of the analyses. In addition to stepwise regression analysis, C&RT, boosted trees and random forest methods, two additional methods, CHAID, and MARS, were also used for development of QSARs for biliary excretion using the procedure explained above for P-gp models. In some C&RT models, the predicted K_i effects were manually incorporated in the models, when they were not picked by C&RT feature selection automatically.

5.3. Results

This chapter will present the results of QSAR development for P-gp binding followed by the QSAR models for biliary secretion that incorporate predicted p-gp binding values as molecular descriptors.

5.3.1. Modelling the P-gp Dissociation Constant (K_i)

P-gp is an important polyspecific transporter protein that can significantly affect the pharmacokinetics of various pharmaceuticals as well as the effectiveness of chemotherapeutics. Due to the major effect of P-gp efflux system in biliary excretion of compounds, it is important to investigate the structural requirements for P-gp binding and predict the binding constants using QSAR. In this investigation, a large dataset of inhibition constant was collated to investigate the development of a universal model for P-gp binding. To help overcome the problem of heterogeneity of the data from various laboratories, that incorporate various substrates at differing concentrations in the design of their experiments, several strategies were implemented. First, the IC_{50} values were converted to K_i values, which is a more comparable measure of inhibitory activity. Secondly, the molecular descriptors of the probe substrates were also used in the analyses and model development process. Third, docking scores from ligand-P-gp docking experiments were incorporated as a molecular descriptor to aid the prediction accuracy of the models. Fourth, the non-linear decision trees and MARS methods were employed that are flexible; therefore, in theory they should be able to deal with more heterogeneous data.

5.3.1.1. P-gp Ligand Docking

Docking energy for all compounds was calculated using MOE software and was used as a molecular descriptor. First in order to verify the docking methodology using MOE software, the geometries of the docked P-gp/QZ59-RRR and P-gp/QZ59-RRR complexes from X-ray crystallography were compared and RMSD between them was calculated. The RMSD value for this structure after

superimposing the docked and co-crystal structures was 0.77; the absolute RMSD range without superposing was 0.89-6.2 for the top 30 poses.

Figure 5.1 shows the 3D structure of P-gp using MOE software. An example substrate can be seen in yellow at the internal cavity corresponding to QZ-RRR binding site.

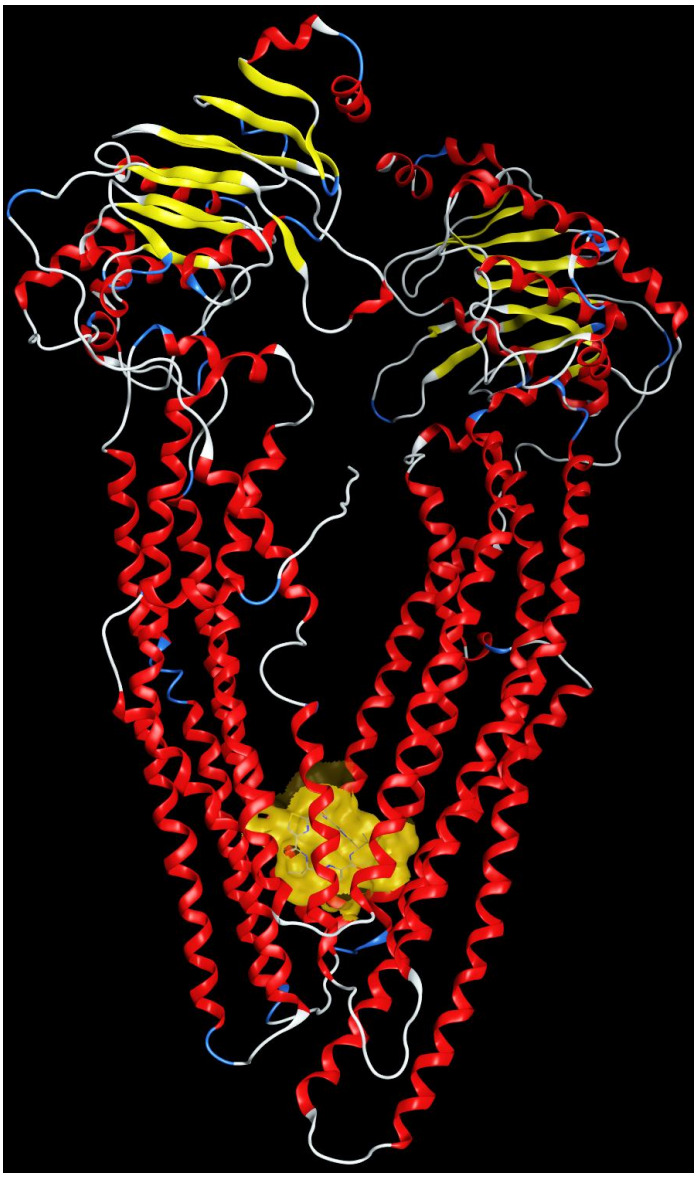

Figure 5.1. Ribbon drawing (front stereo view) of mouse P-gp (PDB id: 3G60) 3D structure in MOE screen shot. The yellow bulb at the lower parts represents the potential binding residues of mouse P-gp in the internal cavity. QZ59-RRR binding site is located in binding pocket in lower side of P-gp cavity. Spiral alpha traces and beta-sheet of P-gp present in red and yellow respectively.

Examples of docking results

Below are examples of P-gp docking of two P-gp substrate/inhibitors namely BMS-387032 (Figure 5.2 and Table 5.1) and SNS-032 (2D diagram is presented in Figure 5.3 and 3D diagram is presented in Figure 5.4). These two compounds have been assessed as potential drugs in multidrug resistant cancer treatment (Michaelis *et al.*, 2014; Löschmann *et al.*, 2013).

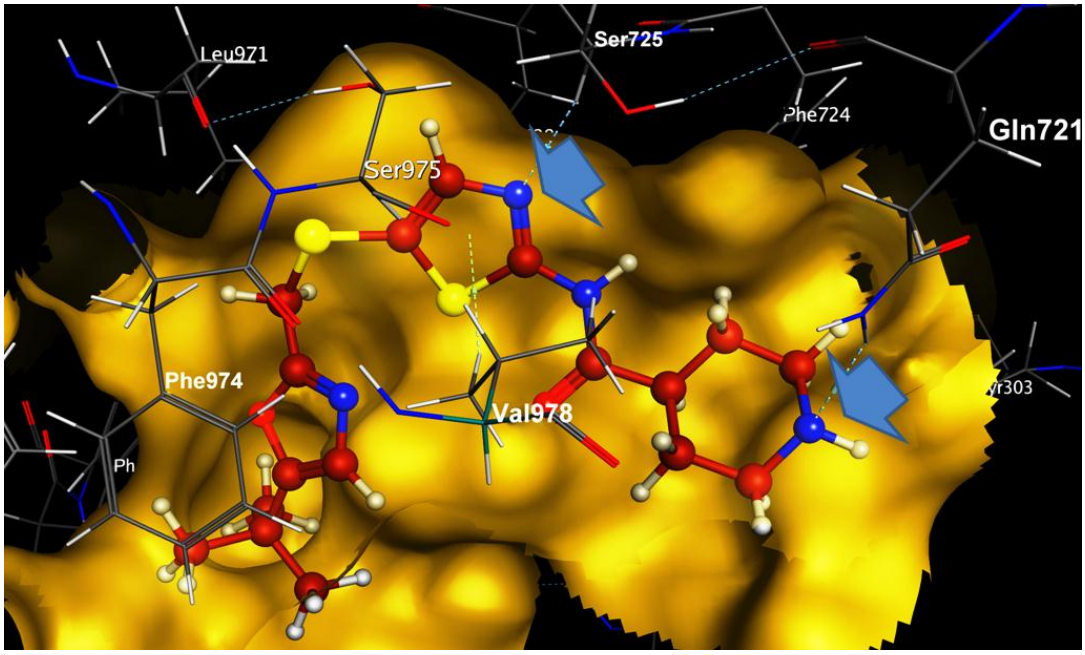

Figure 5.2. The docked conformation of BMS-387032 in the binding pocket of mouse P-gp with the lowest docking energy; blue arrows are strong hydrogen bonds (limited within 4.5 Å) between residues of Ser725 and Gln721 and nitrogen in thiazole and piperidine respectively. Val978 and Phe974 are other residues with pi-H and pi-pi interactions with the BMS-387032 respectively.

Table 5.1. Ligand interactions parameters for binding of BMS-387032 to mouse P-gp (3G60) at the QZ59-RRR binding site (first docking pose)

Fragment of Ligand	Receptor	Interaction	Distance (Å)	E (kcal/mol)
Nitrogen in Thiazole	SER725	H-acceptor	3.47	-0.7
Piperidine	GLN721	H-acceptor	3.10	-1.9
Thiazole	VAL978	pi-H	3.44	-0.9
Thiazole	PHE728	pi-pi	3.92	-0.0
Oxazole	PHE 974	pi-pi	3.68	-0.0

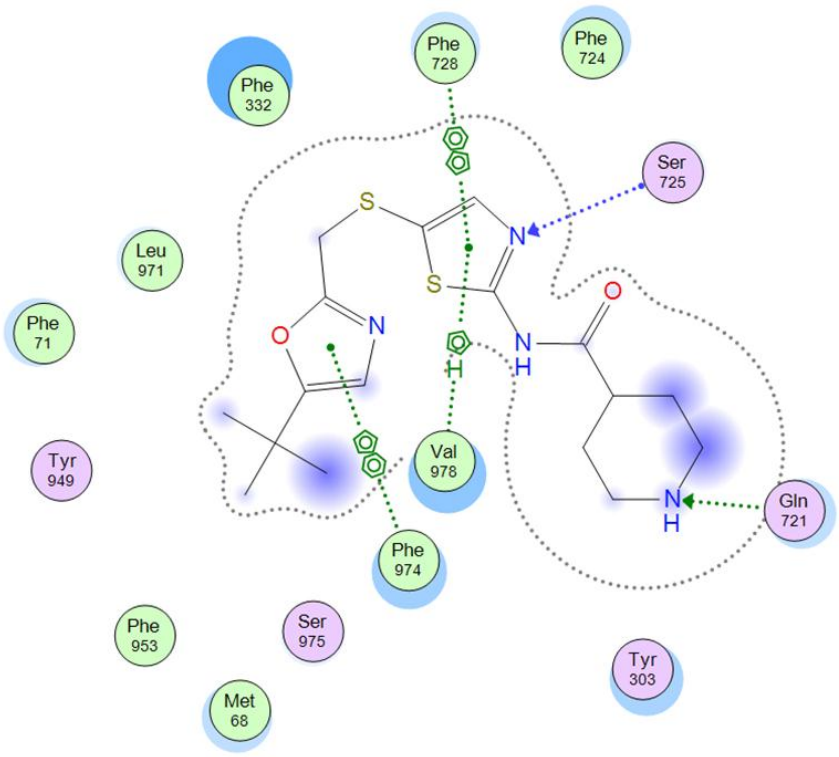

Figure 5.3. 2D graph of interaction of SNS-032 with the QZ59-RRR binding site of P-gp using MOE software; the diagram indicates the polar and non-polar interactions by pink or green coloured amino acids; hydrogen bonding is indicated by green dotted arrows and Pi-H interactions with green dotted line. In this diagram, the energy cut-off for H-bond and ionic interactions were -0.5 kcal/mol and the maximum distance for nonbonded groups was 4.5 Å. Proximity contour are dotted lines surrounding the ligand and indicate the shape of the binding site and available space to the more outward-facing parts of the ligand. Blue shadows in some amino acids indicate the receptor exposure differences by the size and intensity of the quoits discs. The directions of the shadow indicate the directions of the amino acids towards the ligands. The blue clouds around the ligand atoms indicate the solvent exposure.

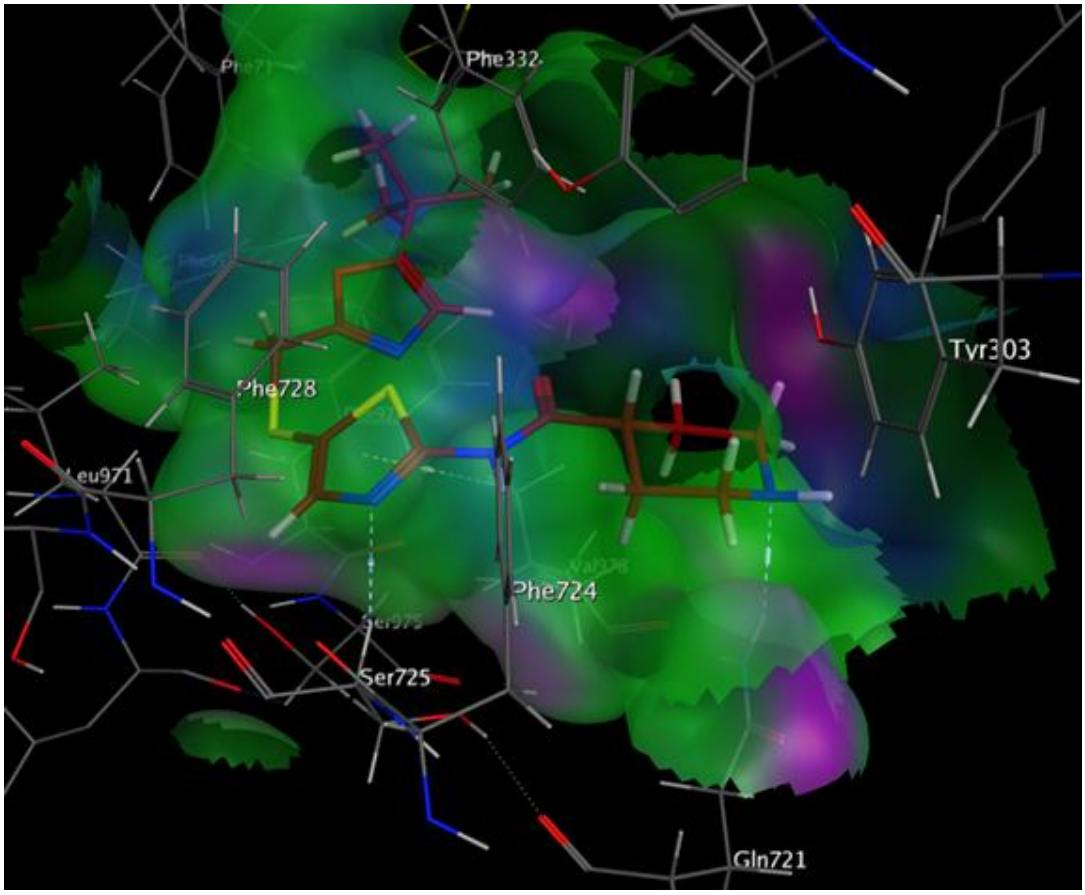

Figure 5.4. 3D diagram of the interaction of SNS-032 with QZ59-RRR binding site of P-gp; the pocket surface is mostly hydrophobic (green colour) and it matches well with hydrophobic rings of the ligand.

8653 poses were obtained after P-gp docking with 219 compounds and the top pose docking energy for each ligand was used as an additional descriptor. The docking study of P-gp inhibitors was carried out using 3D structures of mouse P-gp (Aller *et al.*, 2009).

5.3.1.2. QSAR Models for P-gp Binding

Various decision trees and ensemble models as well as Multivariate Adaptive Regression Splines (MARS) model were developed for the prediction of P-gp inhibition constant. Table 5.2 summarises the selected models developed using various statistical methods. All models obtained are cross-validated and pruned automatically, and the selected models are those with the lowest standard error for

the internal and external test sets. Models listed in Table 5.2 are results of various feature selection and data analysis methods. Majority of these models can be easily interpreted in terms of the molecular characteristics required for an effective P-gp inhibitor. Here we provide a brief description of the models and the inferred molecular characteristics. The molecular descriptors employed in these models have been described in Table 5.3.

Table 5.2. Standard error for the training and internal test sets for the selected P-gp models

Model	Descriptors supplied	Descriptors incorporated manually	Group	Risk Estimate	Standard Error
RT (2)	All descriptors	-	Train	0.246	0.028
			Test	0.810	0.118
CHAID (1)	All descriptors	-	Train	0.420	0.054
			Test	0.672	0.077
I-tree (3)	All descriptors	Docking energies	Train	0.448	0.050
			Test	0.785	0.148
BT (3)	All descriptors	-	Train	0.146	0.013
			Test	0.572	0.126
RF (2)	All descriptors	-	Train	0.438	0.057
			Test	0.607	0.127
MARS (1)	Selected descriptors	-	Train	-	0.048
			Test	-	0.128

Table 5.3. A brief description of the most important molecular descriptors selected and used by the models.

Descriptor	Model	Description
balabanJ	RT (2)	Balaban averaged distance sum connectivity index
b_double	RT (2)	Number of double bonds.
Docking energy (MOE)	I-trees (3)	Docking score (kcal/mol)for enzyme-ligand docking of the compounds into the active site of P-glycoprotein (Aller *et al.*, 2009) calculated using MOE software
GCUT_SMR_2	BT (3)	The GCUT descriptors using atomic contribution to molar refractivity (4 descriptors).
GCUT_SMR_3	MARS (1)	See GCUT_SMR_2.
logP (o/w)	MARS (1)	Log of the octanol/water partition coefficient.
Num Rings 3	CHAID (2)	Number of rings 3

opr_leadlike	CHAID (2)	This is one if and only if there are fewer than two violations from Oprea's lead like rules, otherwise zero
PEOE_VSA+0	RT (2)	van der Waals surface area of atoms with atomic charge in the range [0.00,0.05).
PEOE_VSA_HYD	RT (2), MARS (1)	Total hydrophobic van der Waals surface area
SaaN_acnt	CHAID (2)	Count of all E-states for aromatic nitrogen atoms
SdsCH	CHAID (2)	Sum of all (H-C=) E-State value in molecule.
S-FRB	BT (3)	The number of free rotatable bonds in a substrate.
S-HAcceptors	BT (3)	The number of hydrogen bond acceptors in substrate.
SHBint4_Acnt	CHAID (2)	Sum of H-bond donors and acceptors indexes separated by four skeletal bonds
S-LogD(2)	RT (2)	Logarithm of distribution coefficient D of a substrate between octanol and buffer layers at pH value 2.0.
SlogP	RT (2), CHAID (2), RF (2)	octanol/water partition coefficient
S-logP	MARS (1), I-tree (3)	octanol/water partition coefficient in substrates.
SMR_VSA2	RT (2)	Sum of approximate accessible van der Waals surface area for atoms with atomic contribution to molar refractivity in (0.26,0.35].
SMR_VSA4	MARS (1)	Sum of approximate accessible van der Waals surface area for atoms with atomic contribution to molar refractivity in (0.39,0.44].
S-PSA	BT (3), MARS (1)	The substrate polar surface area.
SssCH2	RT (2)	Count of all CH2 groups E-state values in molecule.
SssS_acnt	CHAID (2)	Count of all sulphur atoms (SssS) E-state values in molecule.
SsssN	BT (3)	Atom-type electrotopological index for tertiary ammonium groups.
Substrate	CHAID (1)	P-gp substrate
vsurf_CW3	RF (2)	Capacity factor is the ratio of the hydrophilic surface over the total molecular surface, calculated at eight different energy levels (from -0.2 to -6.0 kcal/mol)
vsurf_CW4	I-tree (3)	See vsurf_CW3.
vsurf_D2	MARS (1)	Hydrophobic volume at -0.4 kcal/mol
vsurf_D4	RF (2)	Hydrophobic volume at -0.8 kcal/mol
vsurf_D7	RF (2)	Hydrophobic volume at -1.4 kcal/mol
vsurf_D8	RT (2), RF (2)	Hydrophobic volume at -1.6 kcal/mol
vsurf_DW13	I-tree (3)	Contact distances of the lowest hydrophilic energy descriptors (vsurf_EWmin).
vsurf_EWmin2	MARS (1)	Second lowest hydrophilic energy
vsurf_R	RF (2)	The surface rugosity related to hydrophobicity volume of an agent (The smaller the ratio, the larger is the rugosity).

vsurf_W4	CHAID (2)	Hydrophilic volume.

5.3.1.2.1. Regression Trees

Figures 5.5 and 5.6 show the regression trees obtained using RT and CHAID (1) respectively. In the regression trees, N is the number of P-gp inhibitors, Mu is the average and Var is the variance of log K_i in each node. It can be seen in Figure 5.5 of the RT model that the molecular descriptor selected by C&RT algorithm for the first split of the data is SlogP (octanol/water partition coefficient). The tree indicates that compounds with lower lipophilicity than SlogP=3.179 are less potent inhibitors of P-gp with average log K_i of 1.90. This group of compounds (node 2) may be considered as non-inhibitors, although further splitting in the tree indicates a group of compounds with large non-polar surface area (PEOE_VSA+0 > 75.6) and more than three double bonds to be reasonably good inhibitors (node 37). On the other hand, potent inhibitors are very lipophilic (node 3) especially those having a Balaban topological index (balabanJ) of ≤ 0.977. This is in agreement with previous studies that have described LogP as an important parameter in drug binding to P-gp (Lu *et al.*, 2001; Matsson *et al.*, 2009; Wang *et al.*, 2003). The significance of LogP in P-gp inhibition is due to the presence of several lipophilic and aromatic residues in the binding sites of P-gp (Aller *et al.*, 2009). BalabanJ is a highly discriminating topological index which represents the extended connectivity and the shape of molecules (Thakur *et al.*, 2004) and has been shown to be related to properties such as melting point and solubility (Ghafourian and Bozorgi, 2010). This indicates the favourable interaction of certain molecular shapes with P-gp.

Nature of the substrate used for the measurement of IC_{50} and log K_i values has an effect on the measured inhibitory activity, as can be seen from the division of compounds in node 11 according to the substrate's apparent distribution coefficient at pH 2 (S-LogD(2), where S indicates the parameter refers to the substrate). Substrates such as daunomycin and quinidine are basic in nature which will result in very low distribution coefficient at pH 2 (LogD(2) ≤ -1.265). According to the RT model in Figure 5.5, such substrates will result in higher measured IC_{50} and log K_i for the inhibitors. Compounds of high lipophilicity (SlogP > 3.179) may be

potent P-gp inhibitors despite the lipophilic substrates if they contain a large hydrophobic volume at the highest hydrophobic interaction level (vsurf_D8) and a large surface area of non-polar atoms (PEOE_VSA_HYD), especially if they are not more lipophilic than SlogP threshold 5.587 (node 51). In node 13, if the lipophilic volume is not larger than 83.75, then compounds with many –CH2- groups (which may represent less branching) can be reasonable inhibitors (average log K_i of 1.34).

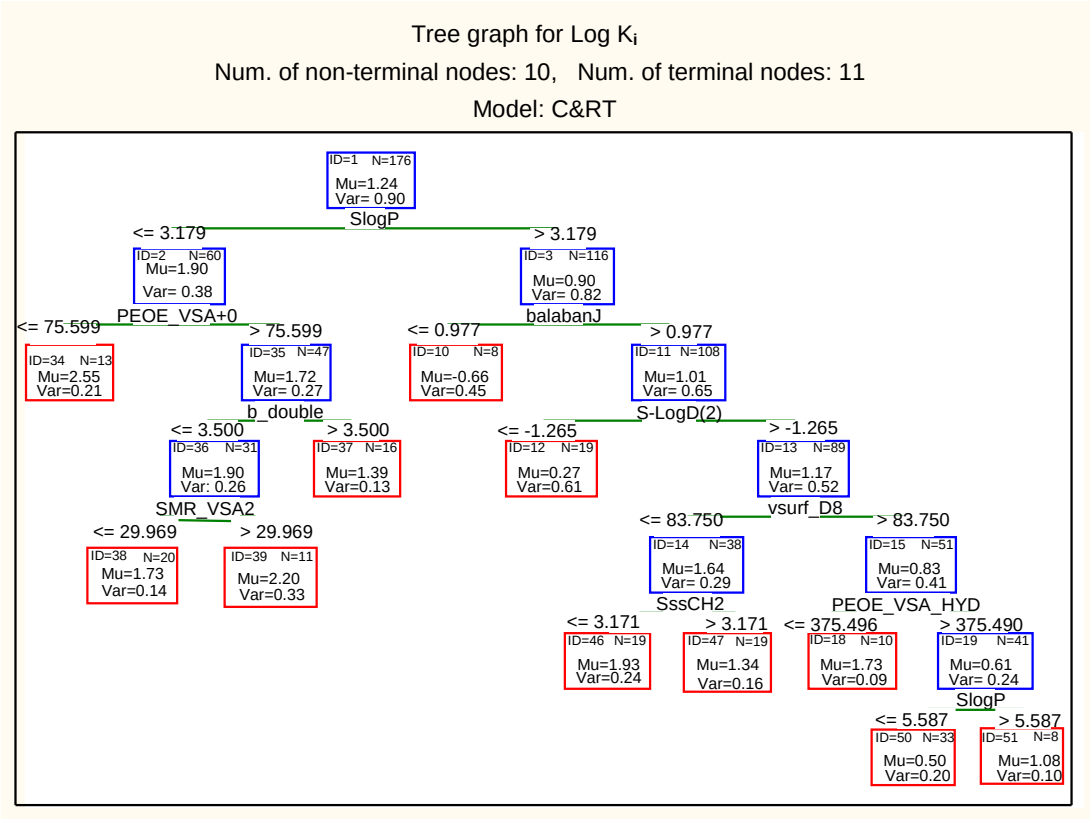

Figure 5.5. RT (2) developed using the training set with the descriptors selected by C&RT algorithm

Figure 5.6 is the selected model developed by CHAID (1) method. Similar to C&RT method above, the hydrophobicity descriptor, SlogP, is the first (most important) descriptor in this CHAID (1) model. In this case compounds have been split into three nodes, with the most lipophilic drugs having the highest inhibition effect (node 4) and the least lipophilic compounds being the least potent or non-

inhibitors (node 2). The non-inhibitors in node 2 have been partitioned further to separate 7 compounds with an aromatic nitrogen group in the structure (SaaN_acnt) as the least effective inhibitors with an average log K_i of 2.78. Node 3 contains compound with intermediate inhibitory activity and SlogP between 2.308 and 3.831. These compounds will be more potent if they contain a double bonding CH group which is seen in compounds such as cyclosporine, valspodar, bromocriptine and quinidine. The most hydrophobic compounds in node 4 are all considered to be strong to moderate inhibitors of P-gp with the log K_i in the terminal nodes ranging from -1.46 to 1.60. In this group, compounds containing 3-membered rings (node 10) and non-lead-like molecules according to Oprea's definition (Oprea, 2000) in node 11 are strong P-gp inhibitors. This observation regarding the higher inhibitory activity of non-lead-like compounds is in agreement with a recent study by Wang et al where lead-like compounds had lower propensity to be P-gp substrates (Wang *et al.*, 2011). Among these inhibitors, those with fewer H-bond donor/ acceptor pairs than two (SHBint4_Acnt) are less strong inhibitors (node 13). In node 13, compounds containing a thioether group are exceptions with a relatively high average log K_i value of 1.60 (SssS_acnt). The remaining 44 compounds (node 17) have high inhibitory activity towards P-gp. Oprea's Lead-like compounds in node 12 may also have strong inhibitory activity towards P-gp if the probe substrate used in the inhibition study is daunomycin.

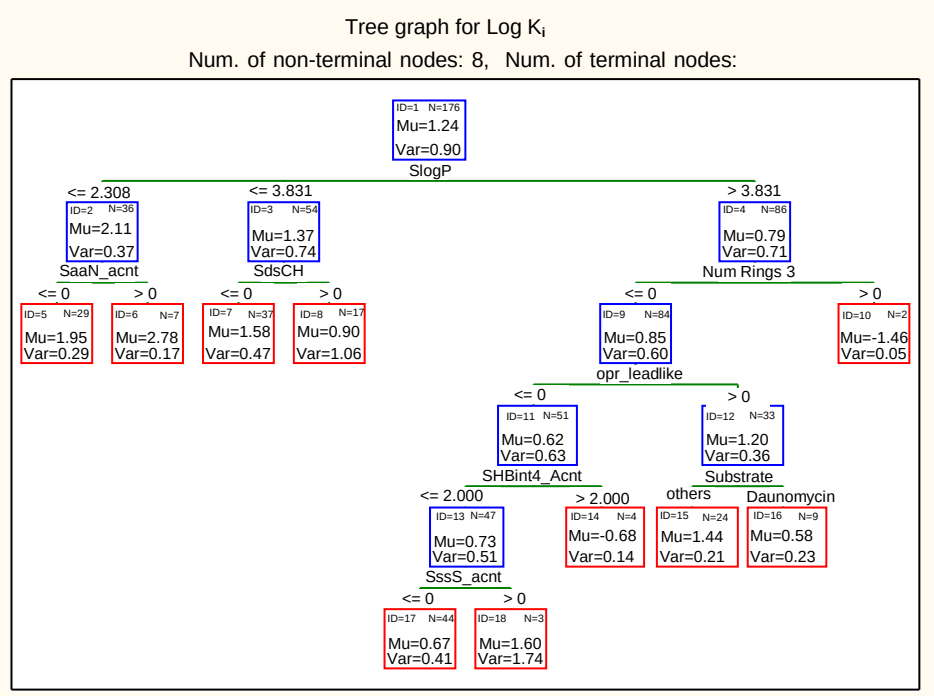

Figure 5.6. CHAID (1) developed using the training set

Despite using P-gp/ inhibitor interaction energies from docking studies as one of the molecular descriptors, none of the decision tree algorithms above, C&RT and CHAID (1), picked docking scores as a significant parameter for partitioning of the log K_i data. This was explored further by using the docking scores in interactive tree, I-tree (3) model (Figure 5.7). In this analysis 'Cross-validate tree sequence' was used in addition to V-fold cross-validation to ensure the validity of each level of the tree for accurate prediction of log K_i in both training and validation sets. Docking score was incorporated as the first variable for partitioning of the data and this was found statistically significant by the cross validations. Figure 5.7 shows that the statistically selected threshold for docking energy is -13.44 (kcal/mol). Inhibitors with docking energy below this value (node 2) will be more effective if they contain a low ratio of hydrophilic to total surface area (vsurf_CW4 $\leq$ 0.539), particularly those with a higher distance between their local hydrophilic energy minima (vsurf_DW13). The tree shows that high docking energy compounds (> -13.44 kcal/mol) are weak inhibitors unless when the probe substrate used in K_i measurement is hydrophilic (S-LogP $\leq$ 0.850).

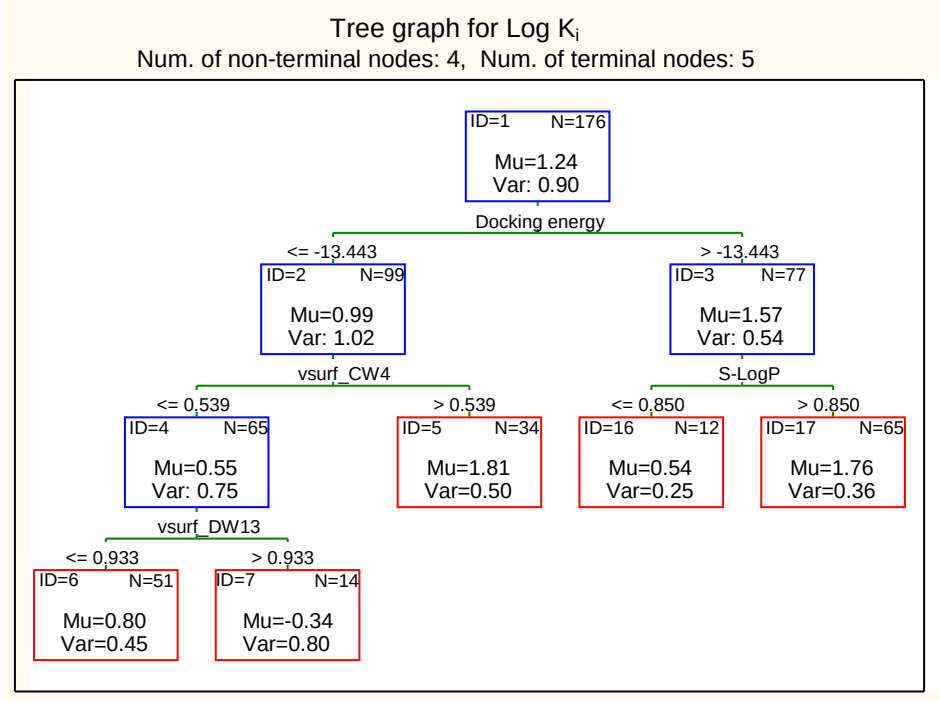

Figure 5.7. I-tree (3) developed using docking energy as the first variable

5.3.1.2.2. Significance of P-gp Docking Energies

Docking is a very useful tool in computer-aided drug discovery due to the importance of shape-matching in drug-macromolecule interactions. It has been postulated that compounds with shape and chemistry similar to those of a known active molecule have a high probability of being active (Hawkins *et al.*, 2007). On the other hand, the interaction energy can be notoriously misleading with large molecular weight compounds often achieving the most negative interaction energies, which is due to the additive nature of the energy formula (Schulz-Gasch and Stahl, 2004; Lipkowitz and Boyd, 2002). In our training set, the top ten molecules with the most negative interaction energies had an average molecular weight of 925 Da in comparison with 461 Da for the remaining compounds in the training set. On the other hand, these ten compounds had a lower average log K_i of 0.75 in comparison with 1.28 for the remaining compounds in the training set.

The lack of flexibility of the target protein during docking should also be taken into consideration when assessing docking results. Docking experiments are most reliable when interaction between a rigid protein target and a flexible ligand is investigated (Davis and Teague, 1999). For docking results to successfully guide

the predictions of inhibitors and substrates of P-gp, it should take into account the very flexible nature of this transporter enzyme (Teague, 2003). Previous studies have described the importance of protein flexibility in P-gp ligand interactions (Loo *et al.*, 2003; Loo *et al.*, 2009). Induced fit mechanism explains the fact that both drug and protein are flexible, and can modify their shape to generate more favourable contacts (Alonso *et al.*, 2006). Current evidence shows that P-gp is able to accommodate a wide range of substrates due to the mobile nature of its transmembrane helices (Loo *et al.*, 2003; Ambudkar *et al.*, 2003). From this hypothesis, it is possible that compounds in the dataset may not be correctly identified as substrates or inhibitors of P-gp, because the docking process does not allow the protein to be mobile and therefore some compounds are not recognised as substrates in the drug binding pocket. Moreover several different but overlapping binding sites have been identified for P-gp (Aller *et al.*, 2009). In this study we used the binding site defined by the cyclic hexapeptide, QZ59-RRR, in the X-ray structure of the protein reported by Aller and co-workers.

5.3.1.2.3. Ensemble Decision Trees

Studies have shown that an ensemble of several trees may result in better prediction accuracy when there is a significant diversity among the models (Kuncheva and Whitaker, 2003). In this investigation boosted trees and random forest were used. Boosted trees method is an ensemble method that computes a sequence of simple trees, each built for the prediction of residuals of the preceding tree. Various combinations of subsample proportions and learning rates were examined and the best model was selected based on the prediction error for the test set. The best result was obtained with the subsample of 0.6 and learning rate of 0.05, using the optimum number of trees of 161. The top ten most important descriptors as calculated by STATISTICA software has been described in Table 5.3. The categorical variable indicating the nature of the substrate was the most important BT (3) descriptor, followed by hydrophobic volume (measured by Volsurf descriptor) and polarity descriptors including total polar van der Waals surface area and total positive and negative partial charges.

Random Forest is another ensemble method which develops a number of decision trees using a random selection of training set compounds and molecular descriptors. The graph of average squared error against number of trees for training and cross-validated test sets indicated that the test error reaches a plateau at around 50-60 trees. Therefore, the final RF model (RF (2)) containing 60 trees was used. In this selected model, descriptors of molecular topology of the inhibitor such as distance and adjacency matrix descriptors as well as lipophilicity indicators and Volsurf molecular interaction descriptors were ranked as the most important descriptors. Unlike the BT (3) model, here there was only one substrate descriptor amongst the top 10 and that ranked as the 10^{th} most important molecular descriptor of the model.

5.3.1.2.4. MARS Model

Many combinations of molecular descriptors picked by several pre-processing feature selection methods were used in MARS analysis to obtain the best possible model as explained in the methods section. The feature selection methods included Chi-square method, stepwise regression analysis, and variable importance rank from random forest and boosted trees analyses. Previous investigations have shown that predictor importance using random forest is a very successful feature selection method that can be applied for reducing the data dimensionality prior to C&RT analysis (Newby *et al.*, 2013a). Here, the best MARS model (Mars (1)) was obtained when the top 15 molecular descriptors from RF model together with the top two substrate descriptors from BT model (S-logP and S-PSA) were given as the independent variables. Subsequently, as a result of the pruning function in MARS analysis, eight out of the 17 molecular descriptors were used in the selected model (summarized in Table 5.4 below). The MARS (1) model in Table 5.4 consists of 11 basis functions with three descriptors employed in two basis functions each and each of the remaining five descriptors are involved in one basis functions. This model does not contain any interaction term. In this MARS model, molecular descriptors have been presented according to the rank order of their importance, with the most important descriptor being the first one in the equation.

An interesting finding from the MARS (1) model in Table 5.4 is a knot at 5.29 for octanol/water partition coefficient, logP(o/w); increasing the lipophilicity of the inhibitors leads to a reduction in log K_i values up to this point. On the other hand, compounds with extremely high lipophilicity (logP(o/w) > 5.29) will have an increased log K_i values (low potency) with increasing their lipophilicity. The second and the third most important descriptors of the MARS model are substrate properties, partition coefficient (S-logP) and polar surface area (S-PSA). Inhibitors will appear less effective (higher measured log K_i values) when the substrate is more lipophilic at S-logP values higher than 2.14. Likewise, substrates of larger polar surface area lead to increased log K_i values. The molecular descriptor derived from the adjacency matrix of the inhibitors (GCUT_SMR_3) is the next most important parameter of the model, which is involved in two basis functions. In this molecular descriptor, the diagonal of the adjacency matrix takes atomic contribution to molar refractivity. The basis functions indicate a positive relationship between log K_i and this molar refractivity indicator for compounds with GCUT_SMR_3 > 3.30; while the opposite (a negative relationship) is observed for compounds having lower molar refractivity indicator. In other words, compounds with high molar refractivity are better inhibitors up to a certain GCUT_SMR_3 threshold. In agreement with this finding, a previous study on P-gp substrates has also indicated a minimum required molar refractivity for the classification of compounds into the substrate category (Demel *et al.*, 2009), but a maximum level of molar refractivity had not been specified. vsurf_D2 is a Volsurf molecular descriptor (Cruciani *et al.*, 2000a), indicating the hydrophobic part of the molecular volume. For the minority compounds with vsurf_D2 < 493 (only 9 compounds), the smaller hydrophobic volumes leads to lower log K_i values. Another molecular descriptor indicating the hydrophobic size of the molecule, (PEOE_VSA_HYD) has appeared in two basis functions with a knot at the descriptor value of 465. For compound with PEOE_VSA_HYD above this threshold value, there is a negative relation with log K_i (the higher the hydrophobic surface area the more effective the inhibitor). The similar trend, but with a much higher gradient, is observed for compounds with PEOE_VSA_HYD < 465. The second lowest hydrophilic energy (vsurf_EWmin2) (Cruciani *et al.*, 2000a), has a negative effect on log K_i, i.e. compounds are less effective inhibitors if the minimum hydrophilic energy is lower than -8.64. This negative impact of a

hydrophilic interaction is only seen for the second hydrophilic region on the molecular surface (not for the first hydrophilic region). Finally, SMR_VSA4 is surface area corresponding to atoms with an atomic contribution to molar refractivity of 0.39-0.44; these are mainly conjugated nitrogen atoms such as those in amide bonds. The MARS equation indicates that presence of more such groups will reduce the log K_i values (better inhibitory effect).

Table 5.4. The selected MARS (1) model

Log K_i = -0.452 + 0.388*max(0, logP(o/w) – 5.29) + 0.255*max(0, 5.29 - logP(o/w)) – 0.475*max(0, 2.14 – S-LogP) + 0.00463*max(0, S-PSA – 45.6) + 3.06*max(0, GCUT_SMR_3 – 3.30) + 0.938*max(0, 3.30 – GCUT_SMR_3) – 0.00684*max(0, 493 – vsurf_D2) – 0.00252*max(0, PEOE_VSA_HYD – 465) + 0.00512*max(0, 465 – PEOE_VSA_HYD) + 0.492*max(0, -8.64 – vsurf_EWmin2) + 0.115*max(0, 3.19 – SMR_VSA4)			
N = 176	GCV error = 0.548	Mean residual = 0.000	SD(residual) = 0.645

5.3.1.2.5. Validation of Models

All models were validated using an external validation set of 43 compounds. Table 5.5 shows the error of the selected models for the prediction of log K_i values of the external validation set and the training set. It can be seen that the RT (2) model gives the most accurate prediction of log K_i followed by BT (3) and then MARS (1). For the training set, BT (3) calculates the most accurate log K_i values followed by RT (2) and then the CHAID (1) model. The difference between model accuracy for training and validation sets may indicate the possibility of overfitting into training data. In this case, amongst the top three models listed above, MARS (1) has the lowest difference between the training and the validation set errors, while BT (3) has the highest difference.

Table 5.5. The summary of the prediction accuracy of the K_i values

Model	MAE for training set	MAE for validation set
RT (2)	0.398	0.543
CHAID (1)	0.471	0.603
I-tree (3)	0.690	0.706
BT (3)	0.316	0.568
RF (2)	0.501	0.618
MARS (1)	0.487	0.577

5.3.2. Prediction of Biliary Excretion Using Predicted P-gp Binding Values

Predicted log K_i by the six models reported in section 5.3.1 were used as independent variables along with the molecular descriptors for the prediction of biliary excretion (log BE%). These were log $K_{i\,(RT)}$, log $K_{i\,(CHAID)}$, log $K_{i\,(I\text{-}tree)}$, log $K_{i\,(BT)}$, log $K_{i\,(RF)}$, and log $K_{i\,(MARS)}$. Models for log BE% were developed using stepwise regression analysis, C&RT, CHAID, boosted trees, random forest and MARS. The results of these analyses have been summarised in Table 5.6. As it can be seen in this table, none of the predicted log K_i values were picked by C&RT, CHAID, stepwise regression analysis (eight parameters), Chi square feature section, MARS feature selection (based on GCV error) or the 20 most important features by random forest, as a significant factor in the estimation of biliary excretion of compounds; the exception to this was the selected BT model. As a result, the multiple linear regression model was the same as MLR (1) (section 4.3.1), and regression trees and random forest models were those reported in section 4.3 (RT (1) and RF (1)).

Table 5.6. Summary of model development for log BE% using molecular descriptors and predicted log K_i values

Method	Predicted log K_i parameter picked	Resulting Model
Stepwise regression	none	MLR (1)
C&RT	none	RT (1)
RF	none	RF (1)
CHAID	none	CHAID (2)
BT	Log $K_{i\ (MARS)}$, Log $K_{i\ (RF)}$	BT (4)
MARS	none	MARS (2)
MARS	Log $K_{i\ (RF)}$	MARS (3)

In this study, in addition to the methods investigated in chapter 4, CHAID and MARS methods were also used for model development. The resulting CHAID model (CHAID (2) in Table 5.6) did not pick any predicted log K_i parameter. This CHAID model has been presented in Figure 5.8.

Figure 5.8 shows that hydrophilic volume (vsurf_W4) is the dominant variable of this tree (node 1), with a binary classification. According to this model, compounds with large hydrophilic volumes are excreted in higher quantities through bile. Other descriptors of CHAID (2) show similar trend to C&RT models presented in Chapter 4 for biliary excretion. For example, hydrophilic compounds with higher acid/base ionisation have higher biliary excretion (node 6), especially if they are non-lead like (node 12). Even compounds with small hydrophilic volumes can have considerable biliry excretion if they are non-lead like (node 4). The high biliary excretion of non-lead-like compounds is in agreement with the results in section 5.3.1 that indicated non lead-like compounds to be suitable P-gp substrates, thereby aiding their excretion by the efflux system. The prediction accuracy of CHAID (2) model is reasonably good (see Table 5.7). The risk estimate and standard error are 0.322 for training set and 0.254 for the validation set.

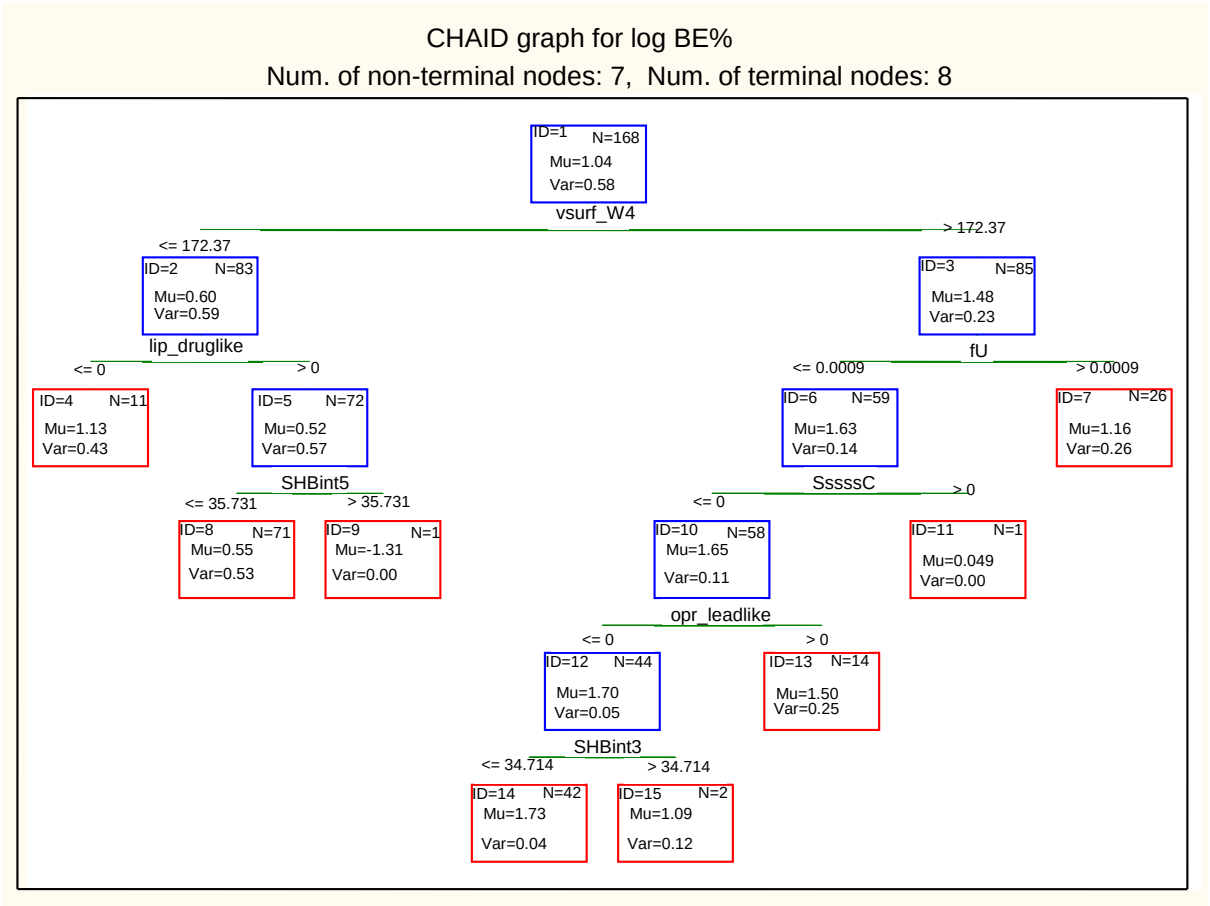

Figure 5.8. CHAID (2) Developed using the training set with the descriptors selected by CHAID algorithm

Table 5.7. Error of biliary excretion (log BE%) prediction by the selected models

Model	MAE for training set	MAE for validation set
BT (4)	0.339	0.416
CHAID (2)	0.432	0.359
MARS (2)	0.438	0.428
MARS (3)	0.436	0.442

As seen in Table 5.6, log K_i predicted by MARS (1) and RF (2) (log $K_{i\,(MARS)}$ and log $K_{i\,(RF)}$) models were two of the most important features in the boosted trees analysis for the prediction of biliary excretion. The selected BT model (BT (4)) has similar prediction accuracy to the BT models without P-gp information (compare BT (1) and BT (2) models in Table 4.5 with BT (4) in Table 5.7). Lipophilicity parameters (LogD (6.5), LogD (7.4)), shape indexes (Kier2, Kier3 and Kier A3)

and Volsurf descriptors indicating hydrophilic ratio (vsurf_CW2 and vsurf_CW4) were amongst the top 15 descriptors of BT (4) model. The optimal number of trees in this graph was 156 (Figure 5.9). Statistical parameters of this boosted tree are reported in Table 5.7.

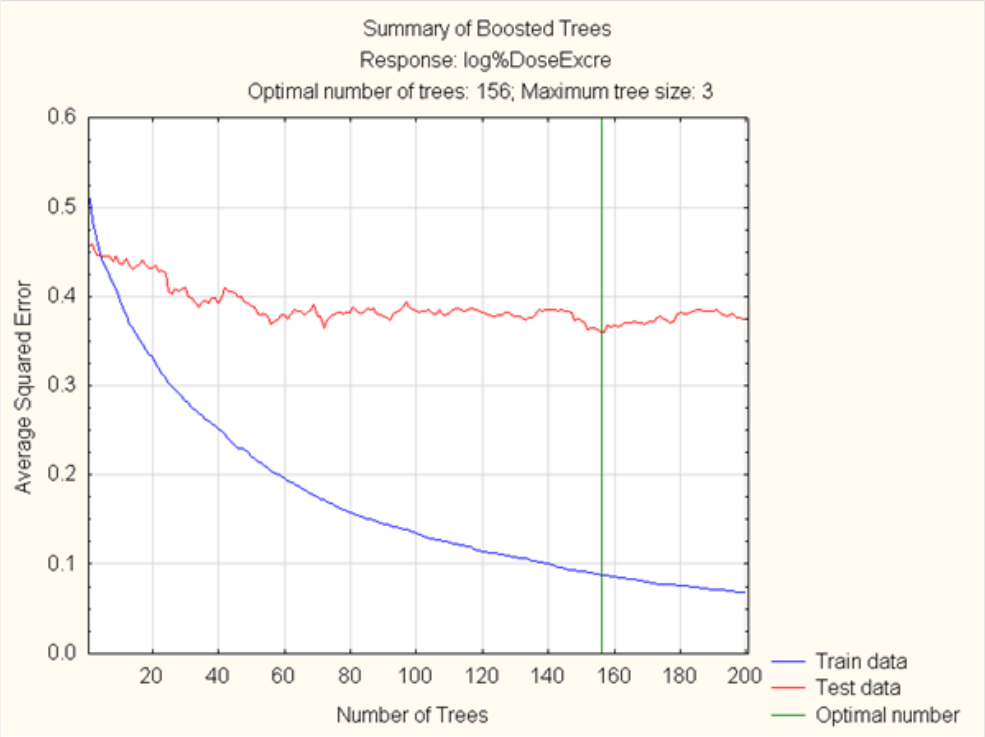

Figure 5.9. Average squared error of log BE% against the number of trees in the boosted trees model BT (4) for the training and internal test sets

MARS models were developed using a number of descriptor sets as explained in the methods section. The best MARS model was MARS (2) using the features selected by Chi square feature method (Table 5.8). The second best model was MARS (3) in which, in addition to Chi square feature predictors, the predicted log K_i values (from RF model) were also used as independent variables. According to MARS (2) and (3), increasing the number of sulphur atoms upto two will increase biliary excretion, with no further increase observed with more sulphore atoms. All the remaining molecular descriptors of MARS (2) are volsurf descriptors of hydrophilic volume and hydrogen bond donor capacity measured at different

energy levels. MARS (3) equation in Table 5.9 indicates that weaker P-gp binders (compounds with higher predicted log K_i values) will have reduced the log BE%. In MARS (3), in addition to the Volsurf (vsurf) variables similar to MARS (2), lipinski's lead-like compounds have been indicated to have lower biliary excretion which is a similar pattern to that observed with P-gp binding.

Table 5.8. The selected MARS (2) model (Feature selection)

Log BE% = -3.14 + 4.99*max(0, vsurf_HB3-8.58) - 3.74*max(0, 9.12-vsurf_W2) + 1.63*max(0, vsurf_W4-1.49) + 3.21*max(0, vsurf_W2-1.24) - 1.99*max(0, 2.00-a_nS) - 1.17*max(0, vsurf_W3-8.07) + 8.547*max(0, 8.07-vsurf_W3) - 1.14*max(0, vsurf_HB4-1.96)			
N = 168	GCV error = 0.398	Mean residual = 0.000	SD(residual) = 0.573

Table 5.9. The selected MARS (3) model (Feature selection and RF predictor)

Log BE% = 8.270- 1.240 (0, vsurf_HB4-2.67) + 2.867*max(0, vsurf_HB3-8.58) + 5.52*max(0, 8.58-vsurf_HB3) - 3.98*max(0, vsurf_W2-9.12) + 6.88*max(0, vsurf_W4-1.49) + 3.33*max(0, vsurf_W2-1.24) - 1.59*max(0, 2.00-a_nS) - 5.70*max(0, log $K_{i\ (RF)}$-1.90) - 3.66*max(0, lip_druglike-0.00)			
N = 168	GCV error = 0.397	Mean residual = 0.000	SD(residual) = 0.565

5.4. Discussion

5.4.1. Structural Determinants of Potent P-gp Inhibitors

Inhibitors of P-gp can be competitive inhibitors that may bind to the substrate binding site, or non-competitive which may bind to other distinct binding sites such

as the ATP-binding site. An investigation that involved docking of multispecific inhibitors into the ATP-binding domain of P-gp has shown that some of the less lipophilic inhibitors can bind to this site, which may contribute to their inhibitory activity (Neuhoff *et al.*, 2000). On the other hand, the more common, lipophilic inhibitors do not interact with the ATP-binding domain of P-gp. Inhibitors from steroid and flavonoid chemotype are examples that may bind to the ATP-binding site (Conseil *et al.*, 1998; Broccatelli *et al.*, 2011). The inhibitors in the training set in this study did not contain any flavonoids but did contain five steroid structures, testosterone, progesterone, spironolactone, digoxin and cortisol. These steroids are also expected to bind to the substrate binding site. For example, studies for several sex-steroid hormones have shown that these are substrates of P-gp mediated transport as well as being a P-gp enzyme inducer (Kim and Benet, 2004) and digoxin is also a known substrate of P-gp as well as acting as an inhibitor (de Lannoy and Silverman, 1992).

From the description of the models outlined above, it can be seen that lipophilicity is the key factor for P-gp inhibition along with the molecular topology and the size of the inhibitors as well as the nature of the substrate probe. In terms of the lipophilicity, a higher partition coefficient than what is recommended for drug-like molecules (based on Lipinski or Oprea's rules) seems to improve the inhibitory activity towards P-gp. According to the best model (RT), the ideal lipophilicity is SlogP value in the range (3.179, 5.587]. A similar pattern can be observed in MARS model where a lipophilicity threshold of 5.29 has been indicated. Previous studies using classification models have found a higher lipophilicity (log P) for multispecific inhibitors of P-gp in comparison with non-inhibitors (Broccatelli *et al.*, 2011; Matsson *et al.*, 2009), although these studies have not specified a maximum lipophilicity threshold. For P-gp substrates, an even higher lipophilicity requirement has been reported in an investigation using a large set of proprietary GSK compounds, i.e. a log P > 4 for the substrate class (Gleeson, 2008).

In addition to the partition coefficient, other lipophilicity measures, which also indicate the size of the lipophilic regions, are found to have an impact. A large hydrophobic volume (vsurf_D8) (Cruciani *et al.*, 2000a), in the RT model and a large hydrophobic surface area (PEOE_VSA_HYD) in MARS and RT models

improve potency of the inhibitors. These two parameters are indicators of both size and lipophilicity. The positive impact of large molecular size and lipophilicity is in agreement with the known structure of P-gp and its proposed substrate binding pocket where the large binding site of P-gp consists of a considerable number of lipophilic amino acids (Song *et al.*, 2010). The descriptor PEOE_VSA_HYD has also been used by Demel et al for the classification of substrates/nonsubstrates, which indicates compounds with PEOE_VSA_HYD > 300, log P < 7 and hydrogen bond acceptor groups more than seven are substrates of P-gp (Demel *et al.*, 2009). Lipophilicity and molecular size have also been indicated in local QSAR models for individual classes of modulators/ substrates (Wang *et al.*, 2003).

In addition, the higher inhibitory activity of non-lead-like compounds (based on Oprea's definition) in CHAID model (CHAID (1)) may also indicate the positive effect of high molecular size and higher lipophilicity than lead-like molecules. Compounds that accommodate the opera's test are defined as compounds with molecular weight $\leqslant$ 460 Da, -4 $\leqslant$ Log P $\leqslant$ 4.2, Log Sw $\geqslant$ -5, number of rotatable bonds $\leqslant$ 10, number of rings $\leqslant$ 4, number of hydrogen donors $\leqslant$ 5, and number of hydrogen acceptors $\leqslant$ 9 (Oprea, 2000). According to this CHAID model, compounds that violate more than two of the above rules are better inhibitors of P-gp. A close observation of such compounds indicates higher lipophilicity or hydrogen bonding groups, as well as higher molecular size and number of rings are the reason for the violations that results in compounds being potent inhibitors. Examples are paclitaxel, nicardipine and vinblastine.

Other significant molecular determinant of P-gp inhibitors is the molecular topology and shape as described by the adjacency and distance matrix descriptors such as the connectivity index BalabanJ in the RT (2), GCUT descriptors in the MARS model and VDistMa in the BT (3). Broccatelli and co-workers (Broccatelli *et al.*, 2011) have also hypothesised that an optimal shape may exist for P-gp inhibitors, but the optimal shape needs to have adequate lipophilicity and H-bond acceptor ability. H-bond acceptor ability has also been emphasised by Demel et al (Demel *et al.*, 2009) which show the importance of a high number or a large surface area of H-bond acceptor groups. In the models presented in this study, the effect of H-bond can be seen in the CHAID (1) where compounds containing more than 2

internal H-bonding are more effective inhibitors. MARS model also indicate the positive impact of presence of conjugated nitrogen groups (e.g. amides). A number of molecular descriptors which may indicate H-bonding effect are present in RF and BT models, including negative charge weighted surface area (CASA-) and partial charge descriptors which are indicators of H-bonding (Dearden and Ghafourian, 1999). It must be noted that these parameters as well as the parameters of Demel *et al.* may also relate to the molecular size as larger molecules are more likely to contain many H-bond groups.

5.4.2. Effect of Substrate on the K_i Measured for the Inhibitors

It has been suggested that there are several binding sites for the molecularly diverse spectrum of P-gp substrates, inhibitors and modulators. For example, using equilibrium and kinetic radioligand binding assays, Martin et al established the presence of at least four distinct interaction sites on P-gp which were able to communicate allosterically (Martin *et al.*, 2000). Moreover, various competitive, cooperative allosteric and anticooperative allosteric interactions are possible between the substrates and the regulators (Lu *et al.*, 2001). As a result, the inhibitory activity measured using different substrates will be different for the same inhibitor (Rautio *et al.*, 2006). The x-ray structure of mouse P-gp with 87% sequence identity to human P-gp has recently been described (Aller *et al.*, 2009). It was found that P-gp can distinguish between different 3D shapes, and that stereoisomers may bind to different binding locations. Given the complexity of the binding locations and modes of inhibition, it has been suggested that a single pharmacophore cannot effectively describe the inhibitors of various P-gp substrates, and therefore, for the inhibition of the transport of different P-gp substrates different pharmacophores have been proposed (Ekins and Erickson, 2002). The modelling strategy in this investigation should be able to deal with the diversity of the binding sites. In particular, molecular descriptors of the substrates were incorporated in the model development in addition to molecular descriptors of inhibitors. Moreover, a categorical variable was implemented in all the decision tree models and ensembles. Regression tree is a powerful data mining tool that is able to select the important features for dividing the data into high or low activity

groups (distinct groups of compounds with high or low average log K_i values). The models described above indicate the importance of substrate in the measured inhibitory activity as all the models contain at least one substrate descriptor selected by the feature selection methods.

The average prediction error separately for the inhibitors of different substrates has calculated. Table 3 gives the average error of log K_i prediction for inhibitors of different substrates using the selected models. The table shows that in average, models predict the inhibitory activity of calcein substrates with the highest accuracy. The rank order of the average prediction error (for the external validation set) from the lowest to the highest is for the inhibitors of calcein, digoxin, vinblastine, daunomycin, irinotecan and quinidine as the probe substrates. The lower average error for a specific substrate's inhibitors may be associated with the number of inhibitors of that substrate in the training set, an indication of which is the number in the validation set shown in Table 5.10.

Table 5.10. Number of inhibitors of different substrates and MAE of log K_i prediction for the validation set

Substrate	n	RT (2)	CHAID (1)	I-tree (3)	RF (2)	BT (3)	MARS (1)
Calcein	14	0.388	0.556	0.609	0.356	0.380	0.300
Daunomycin	4	0.668	0.658	0.869	0.850	0.735	0.735
Digoxin	18	0.574	0.985	0.754	0.601	0.634	0.611
Irinotecan	2	1.005	0.726	1.223	1.365	1.418	1.517
Quinidine	1	1.270	0.033	1.809	1.771	1.249	2.582
Vinblastine	5	0.668	0.934	0.866	0.696	0.435	0.559

5.4.3. Effect of P-gp Binding on Biliary Excretion Models

It can be seen from the results that the use of predicted P-gp binding values did not lead to improved models for biliary excretion, and Log K_i was selected only by the BT (4) and MARS (3) models. However, similarities can be observed between molecular determinants of P-gp binding and biliary excretion. For example, Oprea's lead-like compounds have lower P-gp binding (as seen in CHAID (1) in Figure 5.6)

as well as having lower biliary excretion (CHAID (2) in Figure 5.8). Also, Lipinski's drug like compounds with a similar definition to Oprea's rule show lower biliary excretion according to and MARS (3) in Table 5.9. This may relate to larger MWs observed for both the prominent substrates of P-gp and cholephilic compounds. However, there are also differences in structural requirements for these two biological properties. Lipophilicity is a major contributor to P-gp binding (Gleeson, 2008), which requires even higher log P than drug-like molecules as seen from MARS (1), CHAID (1) and RT (2) models in section 5.3.1. The effect of lipophilicity on biliary excretion is different with large hydrophilic molecules being more prone to biliary excretion as lipophilic compounds go through metabolism instead (Sharifi and Ghafourian, 2014). This result should not be considered as contradictory, as metabolism and biliary excretion are simultaneous processes in hepatocytes and the overall effect is determined by the kinetics. It may be speculated that large lipophilic compounds would be able to be excreted through bile if their metabolism was limited/ slowed down.

In analysing the effect of P-gp binding on the observed *in vivo* biliary excretion levels of compounds one should also consider the fact that P-gp binding data has been obtained from *in vitro* experimentations using different cell cultures. This model may not realistically represent the *in vivo* situation with healthy hepatocytes in their natural liver environments. Moreover, P-gp is only one of the several efflux pumps that operate in hepatocytes.

One possible reason for the 'predicted P-gp binding' not being selected by several feature selection methods could be the poor prediction of P-gp binding for the external (biliary excretion) dataset. Although the prediction accuracy for the external validation set in P-gp binding QSARs have been tested to be satisfactory (Table 5.5), the accuracy of prediction of P-gp binding for biliary excretion cannot be assessed as the experimental values are not available for this dataset. The poor prediction accuracy may happen if the diversity of compounds is different between the two datasets, which may result in the biliary excretion dataset to fall outside the applicability domain of P-gp models. According to Netzeva et al (2005) an applicability domain need to be defined for QSAR models when using for external predictions. In order to investigate this, principle component analysis (PCA) was

performed using all the molecular descriptors. Figure 5.10 show the scores plot of PC1 against PC2. It can be seen in the figure that despite a very good overlap, there are many compounds in BE dataset on the left hand side of the figure which are outside the range of, and further away from, the P-gp dataset.

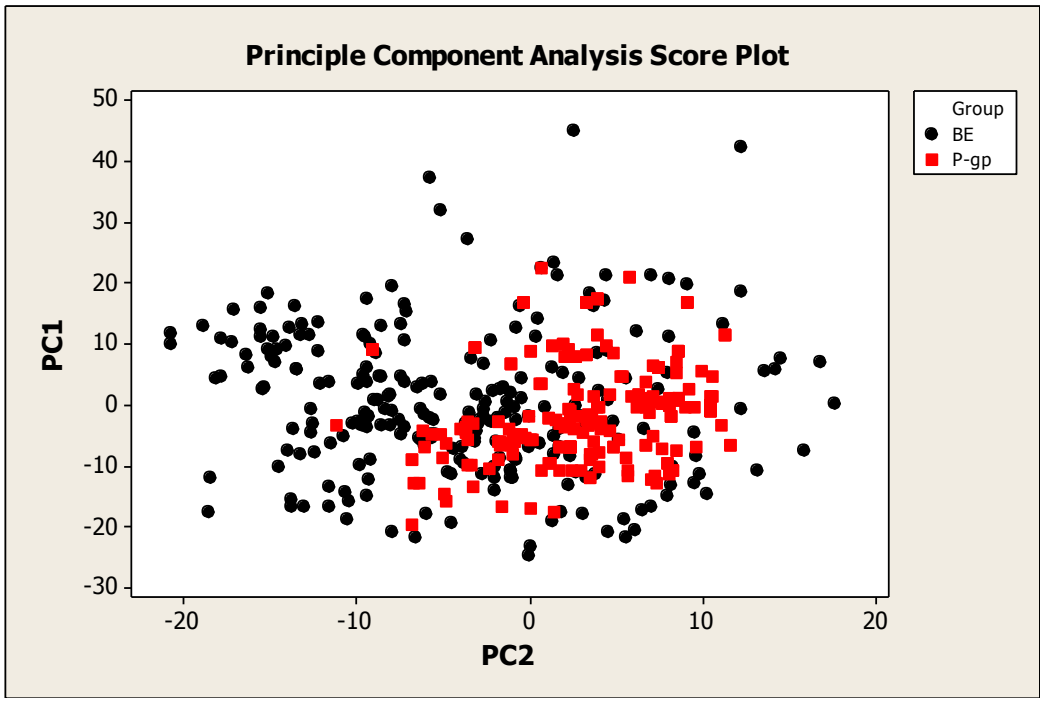

Figure 5.10. Scores plot indicating biliary excretion dataset (BE) and the P-gp binding dataset (P-gp)

5.5. Conclusion

In order to develop accurate models for the P-gp inhibition, this study used K_i values of large set of P-gp inhibitors calculated from the reported IC_{50} and the probe substrate's Km and concentration values from the literature using Cheng and Prusoff's equation. In comparison with IC_{50}, this parameter allows a better comparison between inhibitory activities measured using different probe substrates and substrate concentrations. In addition to the molecular descriptors of the inhibitors, this QSAR study also incorporated the molecular descriptors calculated

for the probe substrate as the nature of the substrate used in the experiment may affect the inhibitory activity of the inhibitor.

The study resulted in a few predictive models based on the accuracy of the prediction for the external validation set. The results indicated that substrate parameters were important for the prediction of the inhibitory activity as all feature selection procedures selected at least one substrate molecular descriptor in addition to the molecular descriptors of the inhibitors. This study also showed that docking scores are not good predictors of inhibitory activity. When used as a molecular descriptor, docking scores were not selected by any of the feature selection methods described here. When docking scores were incorporated manually in C&RT analysis, the resulting regression tree had a high error for the prediction of the validation set. The most significant models indicated a higher lipophilicity of the potent inhibitors than lead-like compounds. The potent inhibitors contained a high molecular weight, a high number/surface area/volume of hydrophobic groups and conjugated nitrogen groups (e.g. amides).

The best model was a regression tree that was obtained using C&RT analysis. A boosted trees model was the second best followed by a MARS equation. Both the regression tree and the MARS model are simple and interpretable and the statistical parameters indicate that they have a lower chance of overfitting in comparison to the boosted trees model.

When the P-gp models were used for the prediction of P-gp binding for the compounds in the biliary excretion dataset, the predicted log K_i values were not picked by several feature selection methods, or when picked (boosted trees and MARS methods), the accuracy of the resulting biliary excretion models were not improved (compare BT (4) with BT (1) and BT (2) or MARS (2) with MARS (3)). This may be attributed to a number of factors including: 1) P-gp is only one of the several efflux pumps operating in hepatocytes, and 2) the poor similarity between the diversity of compounds in the dataset used for P-gp binding models and the biliary excretion dataset may have led to poor prediction of K_i values for compounds in biliary excretion dataset.

6. Inhibitory Effect of OATPs in Biliary Excretion

6.1. Introduction

Several members of the organic anion transporting polypeptide (OATP) family have been shown to be specifically expressed in the liver and facilitate the liver uptake of their substrate drugs. Mechanistic studies suggest an important role for OATP family in the uptake of compounds from blood to hepatocyte, across the basolateral (sinusoidal) membrane (Yamazaki *et al.*, 1996). After transporting the compounds into hepatocytes, these compounds are either metabolised or secreted into the bile using ATP-depandent transporter proteins such as P-gp and MRP2 (Ayrton and Morgan, 2001). In fact, uptake by OATP transporters has often been regarded as the single most important uptake mechanism involved in biliary excretion (Fenner *et al.*, 2012; Varma *et al.*, 2012). For example, studies on lipid-lowering drugs have shown that inhibition of OATP1B1 hepatic uptake can considerably increase statin concentration in blood after administration of cyclosporine, a potent inhibitor of various OATPs (Shitara *et al.*, 2003; Ho *et al.*, 2006), and similar results have been obtained later by Neuvonen and co-workers for other statins (Neuvonen *el al.*, 2006).

Through their role in biliary excretion, OATPs also contribute to drug-drug interaction events (Koenen *et al.*, 2011). As mentioned above, cyclosporin is a potent inhibitor of OATPs (in particular OATP2B1 and OATP1B1) and it is, at the same time, a substrate of CYP3A4, thereby functioning as a competitive inhibitor resulting in increased exposure of other CYP3A4 substrates (Wacher *et al.*, 1998). In addition, this compound interacts with P-gp (Foxwell *et al.*, 1989) and MRP2 (Tang *et al.*, 2002a). These efflux pumps are expressed in the canalicular membrane of hepatocytes. As a result of all these enzyme and transporter interactions, this drug has an impact on the biliary elimination of substrate compounds. Due to the importance of transporters in drug-drug interactions, recently, in drug evaluation process, the identification and kinetic characterization of OATP ligands early on has become important for successful drug development (De Bruyn *et al.*, 2013).

Unfortunately, studies on OATP are limited due to the lack of very specific inhibitor/substrates for this family of transporters. For example, in sinusoidal hepatocyte membrane, apart from OATP1B1 and OATP1B3 which are expressed abundantly, OATP1A2 is also localized in a smaller quantity. All of these three transporters are able to uptake pitavastatin in human hepatocyte. To elucidate which OATP is actually responsible for the pitavastatin uptake, Hirano and colleges investigated the relative contribution of OATP1B1 to the hepatic uptake of pitavastatin. This was done by inhibition of hepatic uptake of pitavastatin by using estradiol-17β-D-glucuronide as an OATP1B1/OATP1B3 inhibitor and estrone-3-sulphate as an OATP1B1/OATP2B1 inhibitor, and comparing their results. The study supported the idea that OATP1B1 is the predominant transporter for the hepatic uptake of pitavastatin (Hirano *et al.*, 2006).

The lack of an X-ray crystal structure is a further limitation with OATP research in the design of the specific modulators. For example, ligand-enzyme docking requires an accurate high-resolution structure of the protein (Rognan, 2013). In a recent investigation, a high-throughput *in vitro* transporter inhibition assay was reported for the OATP1B subfamily (De Bruyn *et al.*, 2013). This approach was able to identify 212 and 139 molecules as inhibitors of OATP1B1 and OATP1B3.

Many OATPs share common substrates. OATP substrates are relatively large from 334 Da in benzylpenicillin to 1143 Da in cholecystokinin octapeptide, in terms of the currently known substrates. Structural templates of many OATP substrates are steroidal or peptidic (You and Morris 2007). The substrate specificity of OATP1B1 is similar to OATP1B3 and both transport a varied range of compounds including bile acids, conjugates of sulphate and glucuronate, steroid conjugates, thyroid hormones, peptides and amphiphilic organic drugs (Glaeser and Kim, 2006; Leuthold *et al.*, 2009; Hagenbuch and Meier, 2003; Tirona *et al.*, 2001; Hsiang et, 1999; Konig *et al.*, 2000a). Many solutes transported by OATPs are negatively charged, however there are several examples of neutral (e.g. digoxin) and cationic (e.g. N-methylquinidine) substrates. Several OATP substrates are promiscuous but there are also some selective substrates. For example, the cholecystokinin octapeptide is a selective OATP1B3 substrate (Nozawa *et al.*, 2003).

The aim of this investigation was to incorporate information from OATP binding in order to improve accuracy of the predicted biliary excretion. This work was carried in two stages: 1) developing the predictive models for OATP inhibition; and 2) using the models for the prediction of OATP effect for the compounds in biliary excretion dataset. OATP models consisted of both regression type (continuous) models and classification type models. Unfortunately, there is a lack of sufficient quantitative data on OATP substrates and non substrates (especially for OATP1B3 and OATP2B1). In a recent study, Varma et al (2012) compared the chemical space of a list of OATP substrates with that of cholephilic compounds. This study suffers from a lack of non-substrate compounds that limits any quantitative conclusion. Karlgren and co-workers (2012a) have recently published a relatively large dataset of OATP inhibition effect measured using high-throughput methods. The measured values are percentage inhibition of a probe substrate's uptake by a large set of compounds. It is noted that a single-point inhibition measure (percentage inhibition) that uses only one inhibitor concentration is not as reliable as IC_{50} for measuring the inhibition activity. Moreover, direct kinetics measures for the substrates would have been the ideal parameter for this investigation. Despite this, considering that most enzyme inhibitors are usually also the substrates of the same enzyme (competitive inhibition), this percentage inhibition dataset was used in this investigation. The single point inhibition assays have proven useful in the past for fast screening of compound activity and selectivity. An example is comparable accuracy of models based on single point CYP inhibition measures, with those built from IC_{50} data (Carlson and Fisher, 2008).

6.2. Methods

6.2.1. Dataset

The dataset of 225 compounds collated, or experimentally determined, by Karlgren and co-workers (2012a) were used in this study. The OATP subfamilies,

OATP1B1, OATP1B3 and OATP2B1 were included in the dataset. A total of 142 compounds in this dataset was from an earlier investigation (Karlgren et al 2012b), which was then expanded to include compounds known to interact with OATPs or CYP enzymes (Karlgren *et al.*, 2012a). The compounds were from the chemical space of oral drugs (Karlgren *et al.*, 2012a). Data consisted of percentage OATP inhibition by the compounds.

The experimental measurements were performed using the human embryotic kidney 293 (HEK293) cells stably transfected with OATP1B1, OATP1B3 or OATP2B1. In the screening experiments to measure interaction of the 225 compounds with each individual OATP, a concentration of 20 µM of the compounds was used. The substrates used in the inhibition studies were estradiol-17β-glucuronide for OATP1B1 and OATP1B3, and estrone-3-sulfate for OATP2B1. The substrate concentration was 0.52 µM in the inhibition of OATP1B1 mediated estradiol-17β-glucuronide uptake. In the inhibition of OATP1B3 mediated estradiol-17β-glucuronide uptake, the substrate concentration was 1.04 µM and in the inhibition of OATP2B1 mediated estrone-3-sulfate uptake, the substrate concentration was 1.02 µM.

The PCA of the dataset indicates that compounds are well distributed in the oral drug space with 95% confidence interval. The dataset included 43% neutral compounds, 29% negatively charged, 22% positively charged and 6% zwitterionic compounds at pH 7.4 (Karlgren *et al.*, 2012a).

For development of QSAR models for OATP interaction, both classification and prediction (regression based) methods were used. The continuous (numerical) percentage inhibition data were used for regression based analyses. For classification methods, compounds were considered as inhibitors if they significantly decreased the uptake of the substrate by at least 50%. In this case, 78 compounds (out of 225 compounds) were OATP1B1 inhibitors, while 46 and 45 compounds (out of 225) were OATP1B3 and OATP2B1 inhibitors, respectively. In the dataset, a few compounds stimulated OATP mediated transporter (instead of inhibition). Clotrimazole, fendiline, progesterone and testosterone are the example of stimulators (Karlgren *et al.*, 2012a). In this investigation all such compounds were considered as non-inhibitors in classification studies.

A total of 387 2D and 3D molecular descriptors were calculated for OATP dataset using the same methods and software as explained in Chapter 4.

6.2.2. QSAR Model Development and Validation

6.2.2.1. OATP Models

Both regression-based and classification models were developed for OATP interaction. The regression based models were linear and non-linear methods of stepwise regression analysis, C&RT, BT, RF and MARS. The classification method was C&RT. All statistical analyses were performed using STATISTICA Data Miner v11 (StatSoft Ltd.).

The compounds were divided into external validation set and training data. Models were developed using training set compounds and assessed using external validation sets. To divide the compounds, they were ordered according to their inhibition percentage and from every set of five compounds, four were allocated into the training and one into the external validation set by random. In this way, training data consisted of 180 compounds and external validation set consisted of 45 compounds. For the analytical methods that required parameter optimization, a fraction of training set compounds were randomly assigned into internal validation set, or alternatively cross validation was used if the option was available in the statistical software. For the internal validation set, where applicable, the risk estimate and standard error were calculated in STATISTICA software and used as the performance indicators.

In OATP modelling using boosted trees, the default values for learning rate, the number of additive terms (number of trees), random test data proportion (fraction of data points in testing pool) and subsample proportion were 0.1, 200, 0.2 and 0.5, respectively. In addition to the default values, various subsample proportions of 0.4, 0.45, 0.50, 0.55 and 0.60 were examined in combination with the learning rates of 0.1 and 0.05. The best OATP models were selected based on the performance indicators for the internal validation set.

6.2.2.2 Biliary Excretion Models

QSAR models were developed for biliary excretion using the dataset and methods explained in Chapter 4. In addition to the molecular descriptors, the OATP effects predicted by the selected models from section 6.2.2.1 were used as the independent variables of the analyses. To this end, the selected OATP models from section 6.2.2.1 were used to predict OATP interaction (percentage inhibition values or inhibitor/non-inhibitor classes) for the compounds in biliary excretion dataset (n = 217). In addition to C&RT method, interactive C&RT was used in which the predicted OATP effects were manually incorporated in the models, when they were not picked by C&RT feature selection automatically.

6.3. Results

It has been cited in the literature that presence of OATPs in the hepatocytes may indicate their significance in biliary excretion process (Matsushima *et al.*, 2005; Pfeifer *et al.*, 2014; Shitara *et al.*, 2013). Binding of 225 compounds to three major sub-family members of hepatic organic anion transporting polypeptides (OATP transporters) were available for this analysis. These sub-families were OATP1B1, OATP1B3 and OATP2B1. The ratios of inhibitors to non-inhibitors were different for each of these three proteins, as can be seen in Table 6.1. A total of 387 molecular descriptors were used for the QSAR model development for the training set consisting 180 compounds. The method of data allocation into training and test sets outlined in the methods section ensured that these sets contained similar ranges of percentage inhibition values. The lipophilicity (LogP by ACD software) was between -4.73 and 8.51 for the training set, and -3.26 and 7.28 for the validation set with similar mean values of 2.43 and 2.58 respectively. Molecular weights of the compounds were between 129-1214 Da for the training set and 94-1202 Da for the validation set, with mean values of 405 and 392 respectively.

Table 6.1. Number of inhibitor/non-inhibitor compounds based in 50% inhibition for each OATP sub-family members

Transporter	Inhibitor	Non-inhibitor	Total
OATP1B1	78	147	225
OATP1B3	46	179	225
OATP2B1	45	180	225

Several QSAR models were developed for each sub-family of OATP transporter using the training set compounds. Based on the prediction error for the validation sets, two QSAR models were selected for the prediction of binding to each OATP for the biliary excretion dataset. Section 6.3.1 gives a brief description of the regression based models, while section 6.3.2 gives description of classification models for OATP interaction. The results of using the predicted OATP effects as the independent variables (descriptors) of biliary excretion models have been presented in section 6.3.3.

6.3.1. Regression Models for Binding to OATP Transporters

Percentage inhibition of OATP transport of a probe substrate by compounds were analysed in this study to develop QSAR models. Distribution of the inhibition data showed normal distributions with 'Skewness' values of 0.163, 0.328 and -3.03; logarithmic transformation of this data led to more skewed data distribution. As a result, QSAR models were developed with percentage inhibition as the dependents variable (non-logarithmic scale). Several QSAR models were developed for each sub-family members of OATP including multiple linear regression analysis, C&RT, boosted trees, random forest, MARS and support vector machine analysis. Two best models for each OATP sub-family based on the lowest error rate in the validation set were selected and are presented below.

6.3.1.1 Selected OATP1B1 Models

Random Forest

A random Forest model was the best model for the estimation of OATP1B1 percentage inhibition values of the external validation set. The selected best RF model was achieved using the number of trees set at 100, a subsample proportion of 0.50, and the random test data proportion of 0.3. Figure 6.1 shows the error reducing as the number of trees increases, and reaching a clear plateau by 100 trees. Prediction accuracy of this model has been presented in Table 6.2 and 6.3. Mean absolute error value for the training and validation sets are ~18 and ~21 respectively. It must be noted here that the errors correspond to the percentage inhibition values in non-logarithmic scale which explains the higher order of the observed error.

The most important descriptor (based on predictor importance in STATISTICA) for this model is VAdjMA, which is a bond count descriptor and defines the number of heavy-heavy bonds in the molecule. The other molecular descriptors, in the top ten important molecular descriptor list, were Chi1, the molecular connectivity index, b_heavy, number of bonds between heavy atoms, SMR_VSA3, the surface area corresponding to atoms with (0.35, 0.39] atomic contribution to molar refractivity, VSA, the total van der Waals surface area, Kier1, molecular shape index, logP calculated by ACD software, and the maximum positive hydrogen atom-level E-state value in a molecule (Hmaxpos).

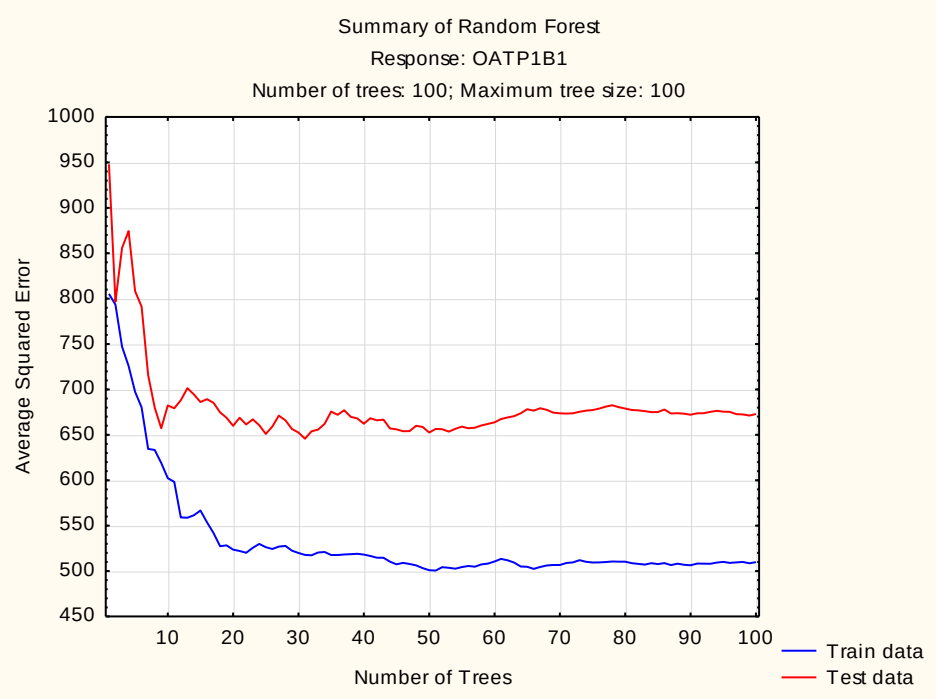

Figure 6.1. OATP1B1-RF model. Average squared error of OATP1B1 against the number of trees in the random forest model (RF) for the training and internal test set

Table 6.2. Statistical parameters of the selected models for training and internal test sets

OATP subfamily	Model	Group	Risk Estimate	Standard Error
OATP1B1	OATP1B1-RF	Train	525	61.1
		Validation	737	135
	OATP1B1-RT	Train	512	58.1
		Validation	690	141
OATP1B3	OATP1B3-BT	Train	487	61.7
		Validation	775	212
	OATP1B3-RF	Train	473	104
		Validation	704	165
OATP2B1	OATP2B1-BT	Train	1959	729
		Validation	1068	239
	OATP2B1-RF	Train	1693	698
		Validation	987	215

Table 6.3. Summary of the prediction accuracy of the selected QSAR models for the training and external validation sets

OATP subfamily	Selected Model	MAE for training set	MAE for validation set
OATP1B1	OATP1B1-RF	17.6	21.0
	OATP1B1-RT	20.6	21.0
OATP1B3	OATP1B3-RF	15.8	20.1
	OATP1B3-BT	16.6	20.3
OATP2B1	OATP2B1-RF	24.3	24.9
	OATP2B1-BT	27.3	25.2

Regression Tree (RT)

The second best QSAR model for OATP1B1 inhibition was a regression tree from C&RT analysis. RT was generated using all molecular descriptors while cross-validation was applied with default V-value of 10 and using interactive C&RT routine STATISTICA. This RT has only one split based on Chi1_C, the carbon valence connectivity index (a topological descriptor). According to this tree, compounds with Chi1_C > 9.698 can bind more strongly to OATP1B1 with an average percentage inhibition of ~68% (node 3). This RT has been presented in Figure 6.2. Table 6.3 shows that despite the very simple nature of this regression tree, the prediction accuracy for the external validation set is similar to the RF model explained earlier.

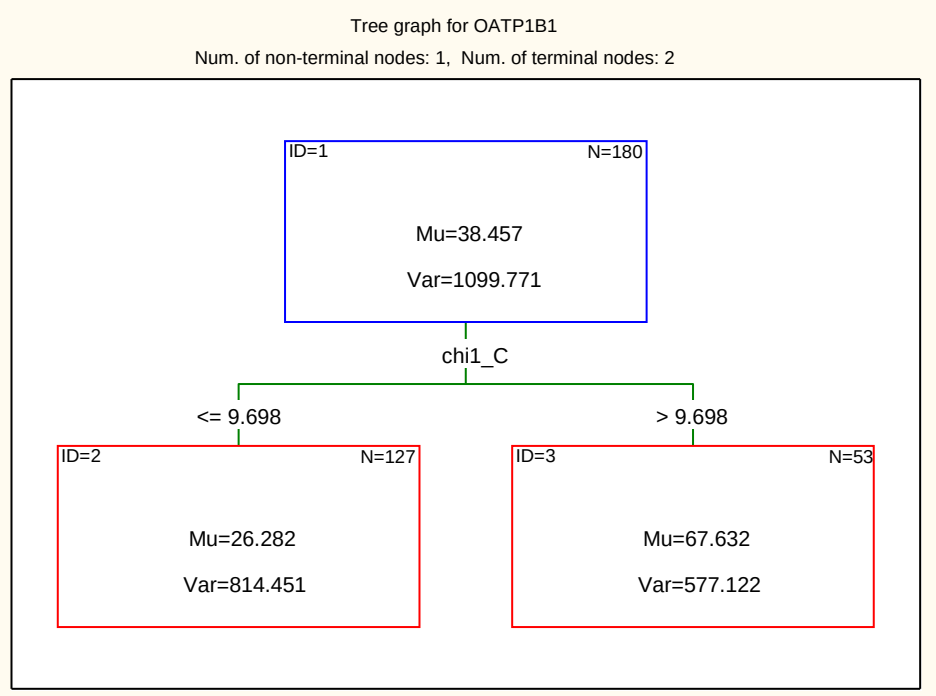

Figure 6.2. The selected RT model for OATP1B1 inhibition developed using C&RT analysis.

6.3.1.2 Selected OATP1B3 Models

Random Forest

The best model for the prediction of OATP1B3 inhibition for the external validation set was achieved using random forest analysis when with a subsample proportion of 0.60 was used and the other statistical parameters were set to default including random test data proportion of 0.3 and the number of trees of 100 (Figure 6.3).

The most important molecular descriptor of the RF model for OATP1B3 is VAdjEq, which is a bond count descriptor and defines the number of heavy-heavy bonds in the molecule. Other most important descriptors of the model are the number of single bonds (b_single), volsurf descriptors indicating hydrogen bonding donor capacity, molecular wrinkled surface and molecular volume (vsurf_HB6, vsurf_R and vsurf_V) and molar refractivity (SMR).

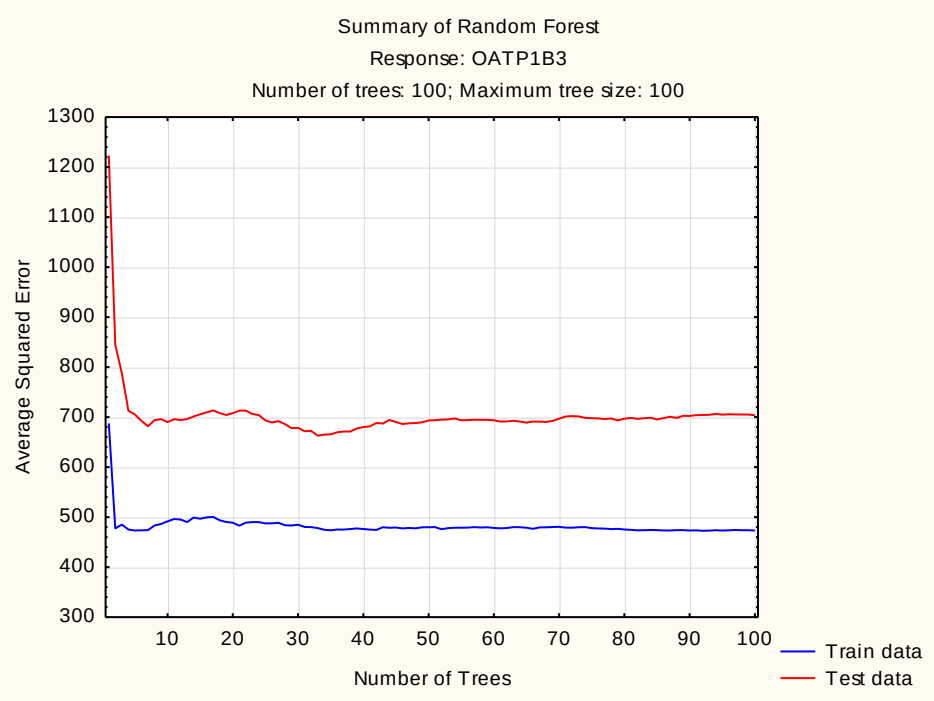

Figure 6.3. Average squared error of prediction of OATP1B3 inhibition against the number of trees in the selected RF model.

Boosted Trees

Boosted trees analysis using various combinations of model parameters resulted in the second best model for the prediction of the OATP1B3 percentage inhibition of the external validation set. In this BT model, the optimal number of trees was 54, with the learning rate of 0.05 and subsample proportions 0.55. Tables 6.2 and 6.3 give a summary of the statistical parameters for the OATP1B3 models. The graph of average squared error against number of trees for training and cross-validated test sets has been presented in Figure 6.4.

The top ranked most significant molecular descriptors of this model in descending order of significance are LogD(10), the apparent partition coefficient at pH 10, FiA, fraction of compound that is ionised as an acid at pH 7.4, SaaCH, atom-type electrotopological index for aromatic CH groups, SaaCH_acnt, the number of aromatic CH groups, the volsurf descriptors, vsurf_IW4 and vsurf_IW5 (indicating hydrophilic integy moments at different levels from -0.2 to 1.6 Kcal/mol),

vsurf_W5 (hydrophilic volume) and SHBint4, internal hydrogen bonding index separated by four skeletal bonds.

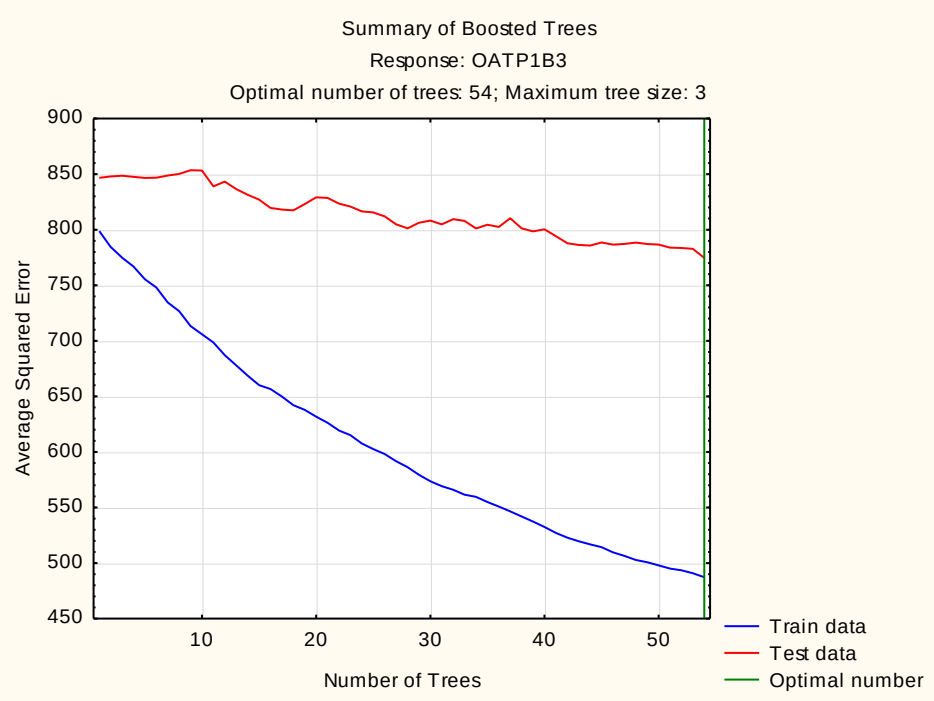

Figure 6.4. Average squared error against the number of trees in the selected BT model for OATP1B3 inhibition.

6.3.1.3 Selected OATP2B1 Models

Random Forest

A Random forest model was the best model for the prediction of OATP2B1 binding of the external validation set compounds. The prediction error for the training and internal test sets as a function of the number of trees has been presented in Figure 6.5. This model was obtained with a subsample proportion of 0.55 and the default parameters of the software. Hmaxpos (the maximum positive hydrogen atom-level E-state value in a molecule) is the most significant molecular descriptor of this selected RF model for OATP2B1. Two BCUT descriptors with atomic contributions to molar refractivity (BCUT_SMR_3) and lipophilicity (BCUT_SLOGP_3), as well as total polar van der Waals surface area (Q_VSA_POL) and fractional negative van der Waals surface area (Q_VSA_FNEG) were the other most important variables of this model.

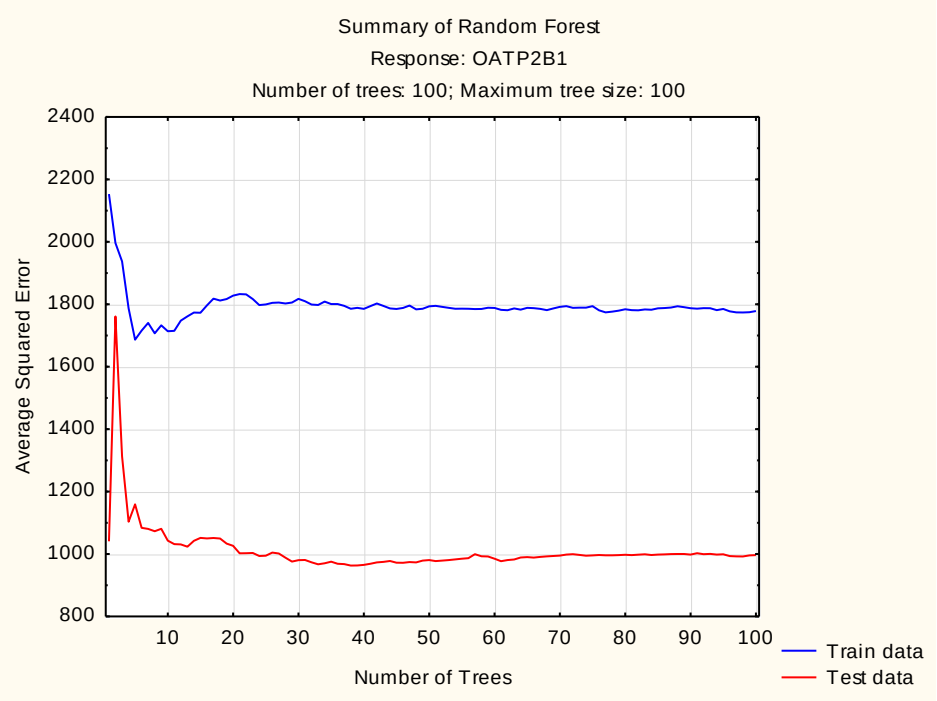

Figure 6.5. Average squared error for the training and internal test sets against the number of trees in the selected RF model for OATP2B1 inhibition.

Boosted Trees

The second best QSAR for the prediction of OATP2B1 binding for the external validation set was obtained using BT analysis when the maximum numbers of trees was 200, with the learning rate of 0.05 and subsample proportions of 0.45 respectively. In the selected BT model the optimum number of trees for predicting OATP2B1 binding of the internal test set was only two (Figure 6.6). Tables 6.2 and 6.3 give a summary of the statistical parameters for the OATP2B1 models.

The most important descriptors using boosted trees analysis were a_ICM, the entropy of the element distribution in the molecule, ratio of carbon atoms in the molecule (C ratio), atom type electrotopological state indexes for various types of carbon atoms (SssssC, SsssCH and SdssC), and the maximum hydrogen atom-level E-state value in a molecule (Hmaxpos and Hmax).

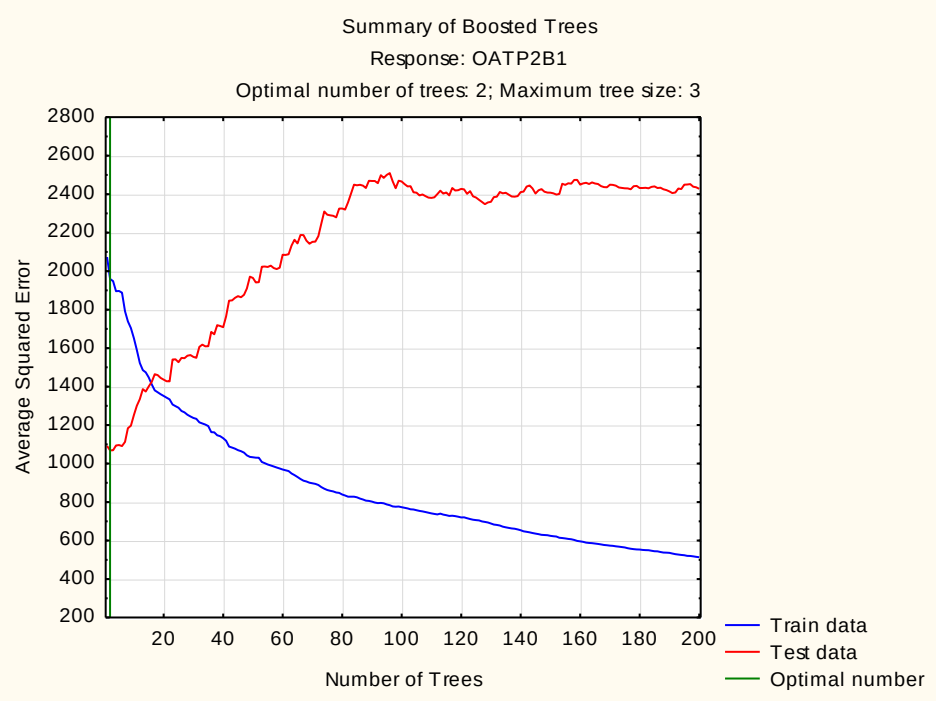

Figure 6.6. Average squared error for the training and internal test sets against the number of trees in the selected BT model for OATP2B1 inhibition.

6.3.2. Classification Models for Binding to OATPs

Due to lower accuracy of percentage inhibition data in comparison with more ideal K_i or IC_{50} data, in addition to prediction (regression) type QSAR models, classification models were also investigated. Classification using C&RT analysis was carried for the dataset of OATP sub-family members. Initially all 387 molecular descriptors were set as independent variables and inhibitor or non-inhibitor class (based on a 50% inhibition threshold) was set as dependent categorical variable. In this way, the classification tree selects the most significant descriptors from the 387 descriptor pool for each split. Figures 6.7, 6.8 and 6.9 show the classification trees for OATP1B1, OATP1B3 and OATP2B1, respectively. Table 6.4 shows the predictive performance measures of the classification trees for OATP models. Sensitivity (SE) shows the percentage of inhibitors predicted correctly and specificity (SP) indicates the percentage of non-inhibitors predicted correctly. Recall that SE, SP and SP × SE should be maximized.

Figure 6.7 shows the classification tree for OATP1B1 binding (CT (1)). Similar to the RT model for OATP1B1 (Figure 6.2), the descriptor chi1_C is the first split variable of CT (1). The cut-off point for the inhibitor class is Chi1_C > 9.68, which is also similar to OATP1B1 RT model. Larger molecules containing many carbon atoms are classed as inhibitors with very few exceptions. An example of exceptions is the compounds with a very low ratio of hydrophilic to lipophilic regions (vsurf_HL1 ≤ 0.05). Compounds classed as non-inhibitor compounds in node 2 are further divided to allow compounds classed as inhibitors if they are very lipophilic (LogD(2) > 4.06), or if they contain an acidic group (partially charged hydrogen atom) (Hmin > 1.39), or if they have a large total negative van der Waals surface area (PEOE_VSA_NEG > 204.43).

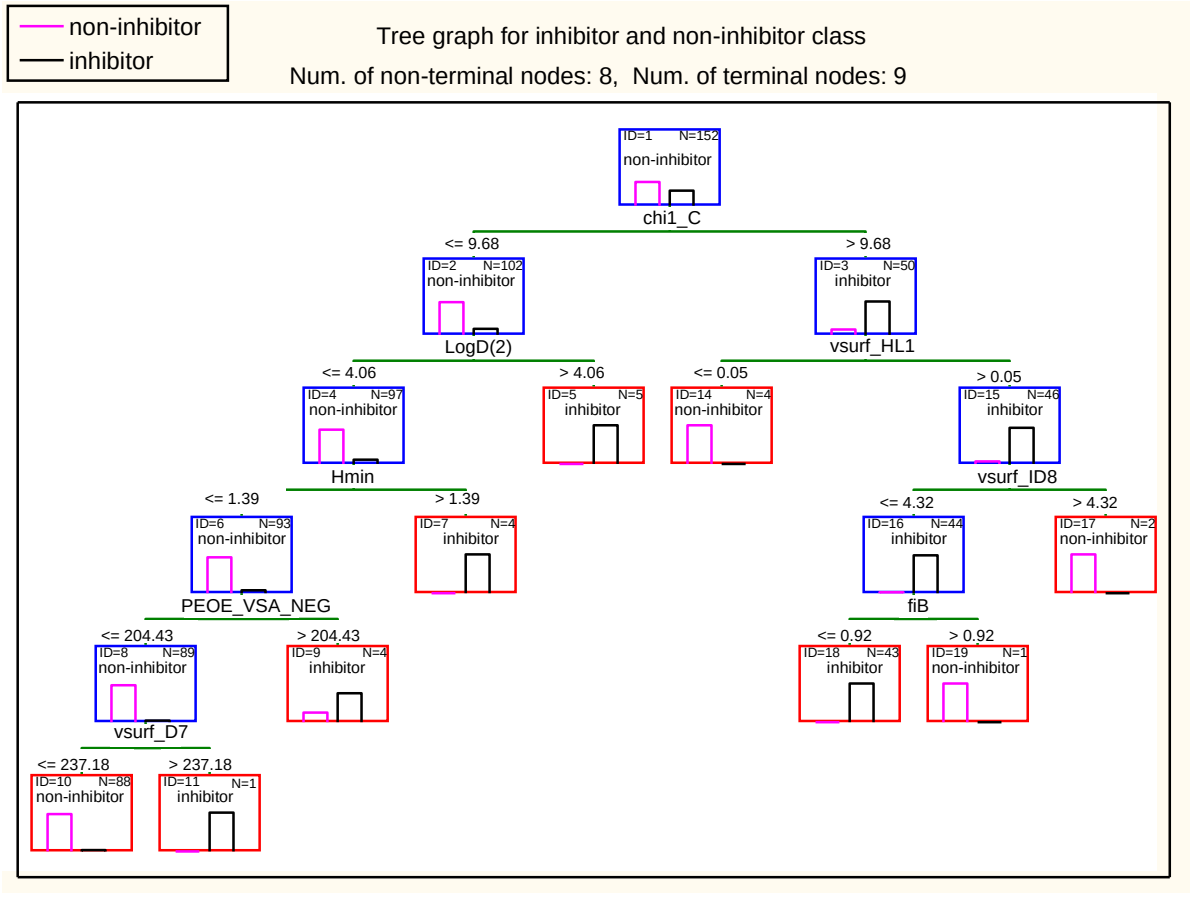

Figure 6.7. CT (1) graph for the best model selecting all descriptors for OATP1B1 50% inhibition

The classification tree for OATP1B3 (CT (2)) is presented in Figure 6.8. The most important molecular property for OATP1B3 inhibitors is a high ratio of rotatable

(single) bonds to total number of bonds in the molecule (b_rotR > 0.3). These flexible molecules need to have a relatively small fraction of polar (to total) surface area to be classed as OATP1B3 inhibitors (Q_VSA_FPOL ≤ 0.36). On the other hand, more rigid molecules can be inhibitors if they have a large total negative polar surface area (Q_VSA_PNEG > 175.05) or a large BCUT_SMR_1 or otherwise, for compounds with large difference between positively charged and negatively charged surface area (DASA), a low BCUT_SMR_2 (≤ 0.067) as well as a low BCUT_SLOGP_1 (≤ -0.47), whereas for compounds with small difference between positively charged and negatively charged surface area, they need a large contact distance between the hydrophilic interaction centres of the molecule (vsurf_DW13).

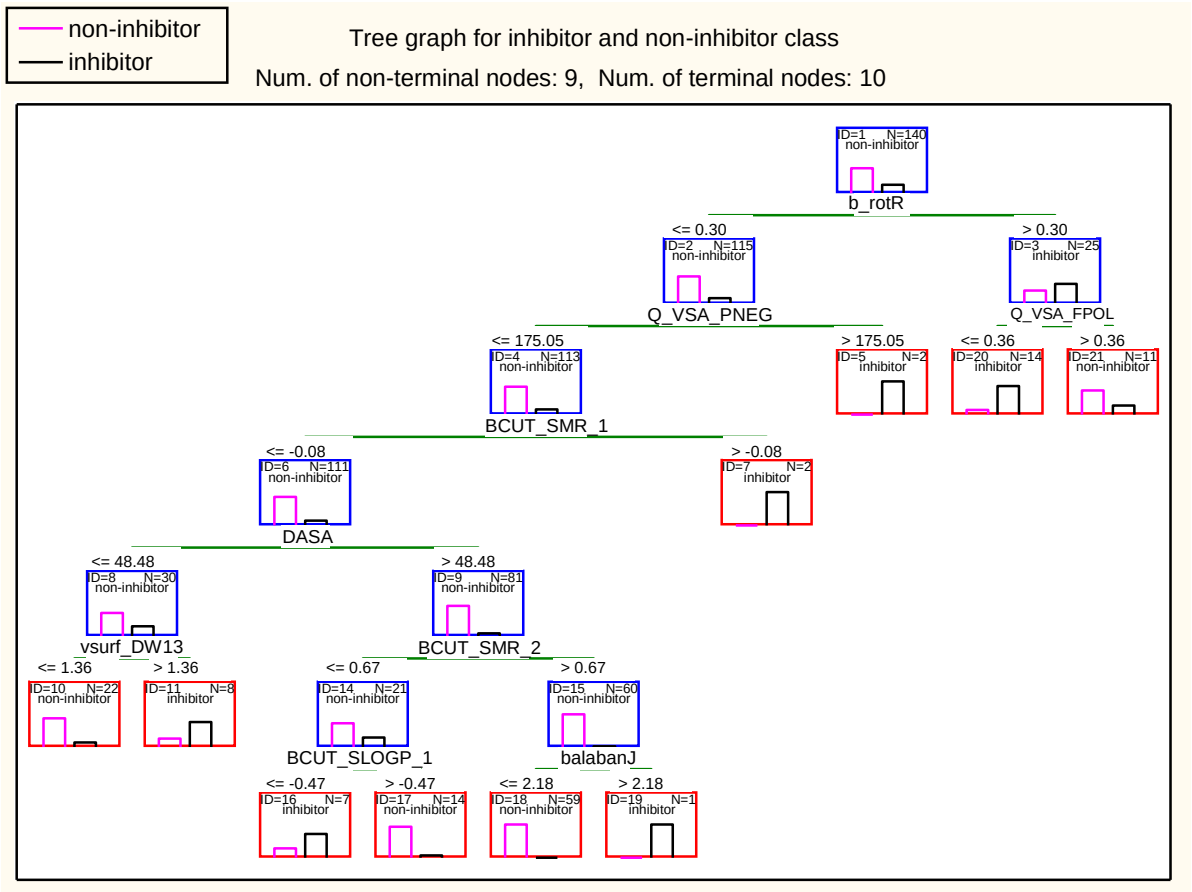

Figure 6.8. CT (2) graph for the best model selecting all descriptors for OATP1B3 50% inhibition

Figure 6.9 shows the classification tree for OATP2B1 (CT (3) model). The first split variable here is vsurf_W1, indicating more hydrophilic drugs (vsurf_W1 >

1412.81) to be inhibitors of this transporter especially if they have a low total positive partial charge calculated by PEOE method (PEOE_PC+ ≤ 4.52), but higher than 4.04 total partial positive charge calculated by AM1 semiempirical method. Less lipophilic compounds will need a GCUT_SLOGP_0 value higher than -0.79 (node 15) to be classed as OATP2B1 inhibitor.

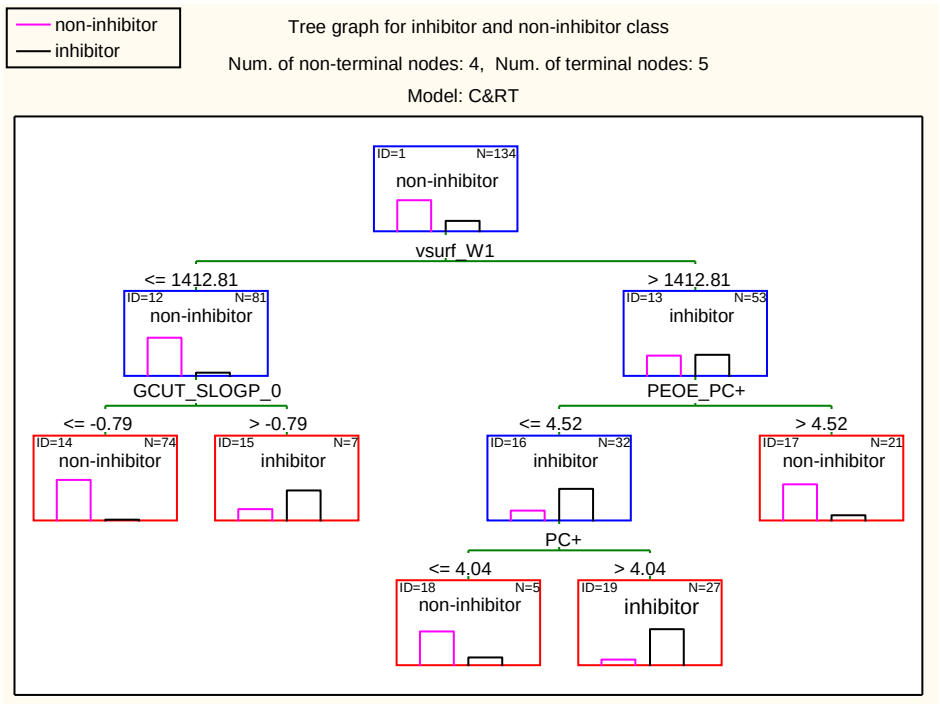

Figure 6.9. CT (3) graph for the best model selecting all descriptors for OATP2B1 50% inhibition

Table 6.4 shows that sensitivity and specificity values are generally good especially for the classification model for OATP1B1 inhibition (CT (1)). All models show better statistics for the training set than for the validation set. The specificity of CT (2) is particularly low for the external validation set. This means that CT (2) cannot classify the non-inhibitors of OATP1B3 accurately, whereas it can predict the inhibitors reasonably well.

Table 6.4. Results of classification analysis using C&RT routines for OATP1B1, OATP1B3 and OATP2B1

OATP subfamily	Model	Set	SP × SE	SE	SP
OATP1B1	CT (1)	Train	0.938	0.989	0.949
		Validation	0.593	0.806	0.736
OATP1B3	CT (2)	Train	0.753	0.942	0.800
		Validation	0.300	0.828	0.363
OATP2B1	CT (3)	Train	0.622	0.882	0.705
		Validation	0.447	0.773	0.578

6.3.3. QSAR Models for Biliary Excretion Using OATP Effects

The selected regression based models from section 6.3.1 were used for the prediction of percentage OATP inhibition by compounds in the biliary excretion dataset. The predicted OATP binding parameters included percentage OATP1B1 inhibition by RF and RT methods (OATP1B1-RF and OATP1B1-RT), percentage OATP1B3 inhibition by RF and BT methods (OATP1B3-RF and OATP1B3-BT) and percentage OATP2B1 inhibition by RF and BT methods (OATP2B1-RF and OATP2B1-BT). These parameters were used as numerical variables in the QSAR model development for biliary excretion of compounds. Moreover, the classification trees from section 6.3.2, CT (1) – CT (3), were used for the prediction of OATP inhibitor/non-inhibitor classes of the compounds in biliary excretion dataset. The predicted classes were used as categorical variable in the QSAR model development using biliary excretion dataset.

6.3.3.1. Regression Tree Models Using Predicted OATP Effects

C&RT analysis was used for the development of a regression tree where log BE% was the dependent continuous variable and the predicted OATP effects along with the molecular descriptors were the independent variables (predictors of the model). The resulting RT (3) model for the training set is presented in Figure 6.10. The molecular descriptors employed in the trees have been explained in Table 6.5.

It can be seen in Figure 6.10 that one of the predicted OATP effects, percentage inhibition of OATP1B3 predicted by RF model (OATP1B3-RF), has been selected by the tree. According to this model, and in agreement with the QSARs discussed earlier (MLR (1), MARS (2) and MARS (3)) for biliary excretion, compounds with large H-bond donor capacity (vsurf_HB3) have higher biliary excretion. The biliary excretion rises further if compounds have high acid/base dissociation (fU $\leq$ 0.001) as seen with previous models such as RT (1) (Figure 4.3). On the other hand, compounds with lower H-bond donor capacity and small negatively charged surface area (Q_VSA_NEG $\leq$ 195.42) are mainly non-inhibitors of OATP1B3 (45 out of 49 compounds node 4) with a low biliary excretion level. Few compounds in node 7 which have been predicted by RF method to be OATP1B3 inhibitors have a very low log BE% (node 7). It must be noted that this result is contradictory to the expectations that compounds with OATP1B3 binding should have more predisposition for biliary excretion. Tables 6.6 and 6.7 provide the statistical parameters of this regression tree, along with all the other models.

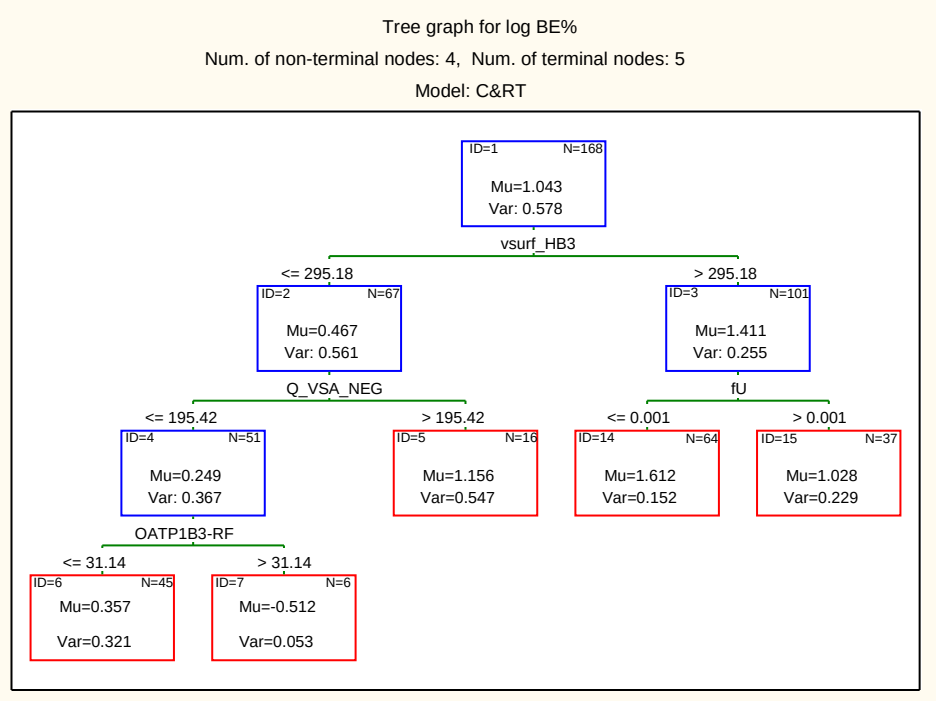

Figure 6.10. RT (3) developed using the training set with the descriptors selected by C&RT

Table 6.5. A brief description of the most important molecular descriptors selected and used by the models.

Descriptor	Model	Description
a_count	RF (3)	Number of atom.
ASA-	RF (3)	Water accessible surface area of all atoms with negative partial charge (strictly less than 0).
balabanJ	CT (2)	Balaban averaged distance sum connectivity index (Balaban, 1982).
b_1rotN	RF (3)	Number of rotatable single bonds (Conjugated single bonds are not included (e.g. ester and peptide bonds)).
b_rotR	CT (2)	Fraction of rotatable single bonds (b_rotN divided by number of bonds between heavy atoms).
b_single	RF (3)	Number of single bonds.
BCUT_PEOE_2	BT (5), I-Tree (9)	The BCUT descriptor (see Table 4.2) using PEOE atomic partial charges.
BCUT_SLOGP_1	CT (2)	The BCUT descriptor using atomic contribution to logP instead of partial charge.
BCUT_SMR_1	CT (2)	This BCUT descriptor using atomic contribution to molar refractivity.
BCUT_SMR_2	CT (2)	This descriptor using atomic contribution to molar refractivity.
chi1v	RF (3)	Atomic valence connectivity index.
chi1_C	CT (1)	Carbon connectivity index.
DASA	CT (2)	Absolute value of the difference between ASA+ and ASA-.
dens	BT (5)	Mass density: molecular weight divided by van der Waals volume as calculated in the vol descriptor.
density	I-Tree (4)	Molecular mass density: Weight divided by vdw_vol (amu/$Å^3$).
fiB	CT (1)	The fractions of compounds ionised at pH 7.4 as base.
fU	RT (3), I-Tree (4), I-Tree (5), I-Tree (6), I-Tree (9)	Fractions of compounds unionised at pH 7.4.
GCUT_PEOE_0	I-Tree (4)	The GCUT descriptors (see Table 4.2) using PEOE atomic charge.
GCUT_PEOE_2	BT (5), I-Tree (6)	See GCUT_PEOE_0
GCUT_SLOGP_0	CT (3)	The GCUT descriptors using the atomic contribution to logP.
GCUT_SLOGP_3	I-Tree (7)	See GCUT_SLOGP_0
glob	I-Tree (9)	Molecular globularity. Globularity or inverse condition number is the smallest eigenvalue divided by the largest eigenvalue of the covariance matrix of

Descriptor	Model	Description
		atomic coordinates. A value of 1 indicates a perfect sphere while a value of 0 indicates a two- or one-dimensional object.
Hmax	BT (5)	Maximum hydrogen E-State atom-level value in a molecule.
Hmaxpos	BT (5)	The maximum positive hydrogen atom-level E-state value in a molecule.
Hmin	I-Tree (9), CT (1)	Minimum hydrogen E-State atom-level value in a molecule.
Kier2	I-Tree (6)	Second order kappa shape index: $(n-1)^2 / m^2$ (Hall *et al.*, 2007).
KierA2	RF (3)	Second order alpha modified shape index: $s\ (s-1)^2 / m^2$ where $s = n + a$
KierFlex	I-Tree (6)	Kier molecular flexibility index: (KierA1) (KierA2) / n (Hall *et al.*, 2007).
LogD(10)	BT (5)	Logarithm of distribution coefficient D of a compound between octanol and buffer layers at pH value 10.
LogD(5.5)	BT (5)	Logarithm of distribution coefficient D of a compound between octanol and buffer layers at pH value 5.5.
LogD(6.5)	BT (5), I-Tree (5)	Logarithm of distribution coefficient D of a compound between octanol and buffer layers at pH value 6.5.
LogD(7.4)	BT (5)	Logarithm of distribution coefficient D of a compound between octanol and buffer layers at pH value 7.4.
LogD(2)	CT (1)	Logarithm of distribution coefficient D of a compound between octanol and buffer layers at pH value 2.
MW	I-Tree (7)	Molecular weight.
OATP1B1-RF	I-Tree (4), I-Tree (7)	Percentage inhibition of OATP1B1predicted by RF model with subsample proportion ratio of 0.50
OATP1B3-RF	RT (3)	Percentage inhibition of OATP1B3 predicted by RF model with subsample proportion ratio of 0.60
OATP1B3-BT	I-Tree (5), I-Tree (7)	Percentage inhibition of OATP1B3 predicted by BT model (with subsample proportion ratio of 0.55 and learning rate of 0.05).
OATP2B1-RF	I-Tree (6)	Percentage inhibition of OATP2B1 predicted by BT model (with subsample proportion ratio of 0.55).
PC+	CT (3)	Total positive partial charge.
PEOE_PC+	RF (3), CT (3)	Total positive partial charge.
PEOE_VSA_HYD	I-Tree (5)	Total hydrophobic van der Waals surface area. This is the sum of the van der Waals surface area such that absolute value of atomic charge is less than or equal to 0.2.
PEOE_VSA_NEG	CT (1)	Total negative van der Waals surface area.

Descriptor	Model	Description
PEOE_VSA+0	I-Tree (5)	Van der Waals surface area of atoms with atomic charge in the range [0.00,0.05).
PEOE_VSA-0	I-Tree (9)	Van der Waals surface area of atoms with atomic charge in the range [-0.05,0.00).
PEOE_VSA+4	RF (3), I-Tree (10)	Van der Waals surface area of atoms with atomic charge in the range [0.20,0.25).
Predicted OATP1B1 Class	I-Tree (8)	This is a categorical descriptor (0 and 1) shows percentage inhibition of OATP1B1 predicted by C&RT routine model.
Predicted OATP1B3 Class	I-Tree (9)	This is a categorical descriptor (0 and 1) shows percentage inhibition of OATP1B3 predicted by C&RT routine model.
Predicted OATP2B1 Class	I-Tree (10)	This is a categorical descriptor (0 and 1) shows percentage inhibition of OATP2B1 predicted by C&RT routine model.
Q_VSA_FPOL	CT (2)	Fractional polar van der Waals surface area. This is the sum of the van der Waals surface area such that absolute value of atomic charge is greater than 0.2 divided by the total surface area.
Q_VSA_PNEG	CT (2)	Total negative polar van der Waals surface area. This is the sum of the van der Waals surface area such that absolute value of atomic charge is less than -0.2.
Q_VSA_NEG	RT (3), I-Tree (5)	Total polar negative van der Waals surface area. This is the sum of the van der Waals surface area such that absolute value of atomic charge is greater than 0.2.
SMR_VSA2	I-Tree (7)	Sum of approximate accessible van der Waals surface area for atoms with atomic contribution to molar refractivity in (0.26, 0.35].
vdw_area	I-Tree (4)	The van der Waals surface area ($Å^2$) calculated using a connection table approximation.
vsurf_D7	CT (1)	Hydrophobic volume (8 descriptors).
vsurf_ID8	CT (1)	Hydrophobic integy moment (The "integy moment" is defined in analogy to the dipole moment and describes the distance of the centre of mass to the barycenter of hydrophobic regions). Small integy moment indicates that the hydrophobic moieties are either close to the centre of mass or they balance at opposite ends of the molecule, so that their resulting barycentre is close to the centre of the molecule. VolSurf computes ID at eight different energy levels (from -0.2 to 1.6 Kcal/mol).
vsurf_CP	I-Tree (6), I-Tree (9)	Critical packing parameter. This parameter defines a ratio between the lipophilic and hydrophilic part of a molecule. It is defined as: volume (lipophilic part)/[(surface(hydrophilic part)(length of lipophilic part)]. Therefore, critical packing refers to

Descriptor	Model	Description
		molecular shape as well as lipophilicity/ hydrophilicity ratio.
vsurf_CW2	BT (5)	Capacity factor is the ratio of the hydrophilic surface over the total molecular surface, calculated at eight different energy levels (from -0.2 to -6.0 kcal/mol).
vsurf_CW4	I-Tree (4), I-Tree (7), I-Tree (6)	See vsurf_CW2.
vsurf_DW13	CT (2)	Contact distances of the lowest hydrophilic energy descriptors (vsurf_EWmin) (3 descriptors).
vsurf_EDmin3	I-Tree (6)	The lowest hydrophobic energy.
vsurf_HB1	RF (3)	H-bond donor capacity at -2.0 Kcal/mol with carbonyl oxygen probe (8 descriptors).
vsurf_HB3	RT (3), I-Tree (5), I-Tree (9), I-Tree (10)	H-bond donor capacity at -2.0 Kcal/mol with carbonyl oxygen probe (8 descriptors).
vsurf_HB4	I-Tree (7)	See vsurf_HB3.
vsurf_HL1	I-Tree (7), CT (1)	Hydrophilic-Lipophilic balance; it is the ratio between the hydrophilic regions measured at -3 and -4 kcal/mol and the hydrophobic regions measured at -0.6 and -0.8 kcal/mol. The balance describes which effect dominates in the molecule, or if they are roughly equally balanced.
vsurf_W1	CT (3)	Hydrophilic volume.
vsurf_W3	I-Tree (8)	Hydrophilic volume.
vsurf_W4	RF (3)	Hydrophilic volume.

Table 6.6. Statistical parameters of the models for training and test sets

Model	Group	Risk Estimate	Standard Error
RT (3)	Train	0.107	0.031
	Validation	0.583	0.118
I-Tree (4)	Train	0.211	0.041
	Validation	0.242	0.053
I-Tree (5)	Train	0.201	0.026
	Validation	0.341	0.087
I-Tree (6)	Train	0.177	0.021
	Validation	0.365	0.086
I-Tree (7)	Train	0.213	0.020
	Validation	0.268	0.069
I-Tree (8)	Train	0.210	0.055
	Validation	0.380	0.067
I-Tree (9)	Train	0.188	0.033

Model	Group	Risk Estimate	Standard Error
	Validation	0.360	0.096
I-Tree (10)	Train	0.247	0.039
	Validation	0.366	0.088
BT (5)	Train	0.087	0.008
	Validation	0.267	0.085
RF (3)	Train	0.280	0.043
	Validation	0.267	0.066

Table 6.7. Summary of the prediction accuracy of the RT models

Model	MAE for training set	MAE for validation set
RT (3)	0.236	0.420
I-Tree (4)	0.343	0.379
I-Tree (5)	0.335	0.409
I-Tree (6)	0.332	0.443
I-Tree (7)	0.362	0.392
I-Tree (8)	0.454	0.455
I-Tree (9)	0.334	0.446
I-Tree (10)	0.448	0.474
BT (5)	0.242	0.362
RF (3)	0.387	0.411

6.3.3.2. Interactive Tree Models Using Predicted OATP Effects

Interactive C&RT analysis was used here to inspect the effect of OATPs more closely. In these analyses one of the most accurately predicted OATP binding (percentage inhibition) or the predicted OATP class was manually used as the first variable in the regression trees for the biliary excretion, and then the tree was allowed to grow automatically using the features selected by the analysis. Hence, we examine the significance of OATPs, namely OATP1B1, OATP1B3 and OATP2B1 in biliary excretion using I-tree analysis. Table 6.8 describes summary of I-tree models in terms of the type of the predicted OATP effect in the model.

Table 6.8. Brief description of the interactive C&RT models

Model no	Manually incorporated variables
I-Tree (4)	Predicted percentage OATP1B1 inhibition using OATP1B1-RF model
I-Tree (5)	Predicted percentage OATP1B3 inhibition using OATP1B3-BT model
I-Tree (6)	Predicted percentage OATP2B1 inhibition using OATP2B1-RF model
I-Tree (7)	Predicted percentage OATP1B1 and OATP1B3 inhibitions using OATP1B1-RF and OATP1B3-BT models
I-Tree (8)	Predicted OATP1B1 inhibitor/non-inhibitor class using CT (1)
I-Tree (9)	Predicted OATP1B3 inhibitor/non-inhibitor class using CT (2)
I-Tree (10)	Predicted OATP2B1 inhibitor/non-inhibitor class using CT (3)

I-Tree (4) (Figure 6.11) shows that compounds with high OATP1B1 binding, as predicted by OATP1B1-RF, have higher biliary excretion. The statistically selected OATP1B1-RF threshold is 37.12. Literally, compounds in biliary excretion dataset that have been predicted to inhibit OATP1B1 by > 37.12% (representing stronger binding to the transporter), are predicted by this model to have higher biliary excretion. Exceptions to this are compounds in node 13, with low hydrophilic surface ratio and high GCUT_PEOE_0. According to this tree, log BE% is low for the non-inhibitors of OATP1B1 with a small van der Waals surface area (vdw_area $\leq$ 297.08) and especially if they have GCUT_PEOE_0 values below -0.85 (node 9). Tables 6.6 and 6.7 provide the statistical parameters of the interactive regression trees.

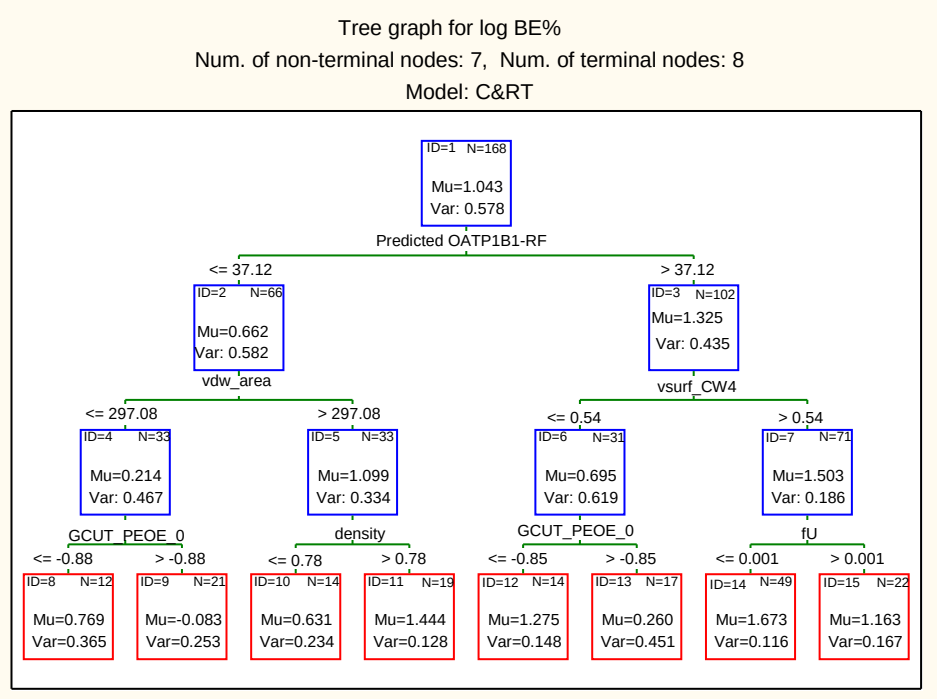

Figure 6.11. I-Tree (4) developed using interactive C&RT analysis using OATP1B1 descriptor as the first descriptor

When predicted OATP1B3 effect (OATP1B3-BT) was used in the analysis, I-Tree (5) was obtained which has been presented in Figure 6.12. According to this tree, 28 OATP1B3 inhibitors (> 52.10% inhibition) have a slightly lower average log BE%. This is due to the effect of 8 compounds in this group with low total hydrophobic surface area (PEOE_VSA_HYD $\leq$ 254.04), which have extremely low biliary excretion (node 6). For OATP1B3 non-inhibitor compounds, log BE% is moderate to high if they have a high H-bond donor capacity (vsurf_HB3 > 298.22) (terminal nodes 10, 14 and 15) or alternatively if they have a large negatively charged surface area (Q_VSA_NEG > 200.31).

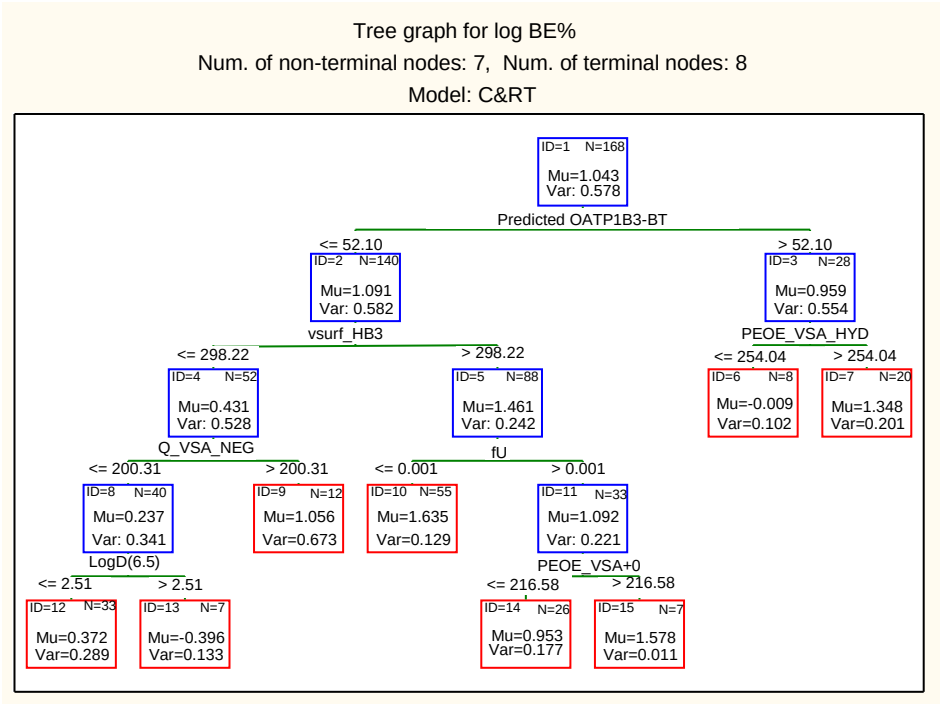

Figure 6.12. I-Tree (5) developed using interactive C&RT analysis using OATP1B3 descriptor as the first descriptor.

Figure 6.13 presents the regression tree using predicted OATP2B1 effect (OATP2B1-RF) as the first split variable (I-tree (6)). The predicted percentage of OATP2B1 inhibition by RF method for compounds in biliary excretion dataset ranged from -1 to 28%. According to this tree, compounds with percentage inhibition above 22.05 have generally higher biliary excretion, except when the compounds are extremely weak acid or bases (fU > 0.001 at pH 7.4) and in addition to their large lipophilic to hydrophilic region ratio (vsurf_Cp > 0.13). On the other hand, OATB2B1 non-inhibitors are generally less excreted through bile, unless if they are large (Kier2 > 8.26) especially if they have GCUT_PEOE_2 > 0.06 (node 11). Statistical parameters of the model can be seen in Tables 6.6 and 6.7.

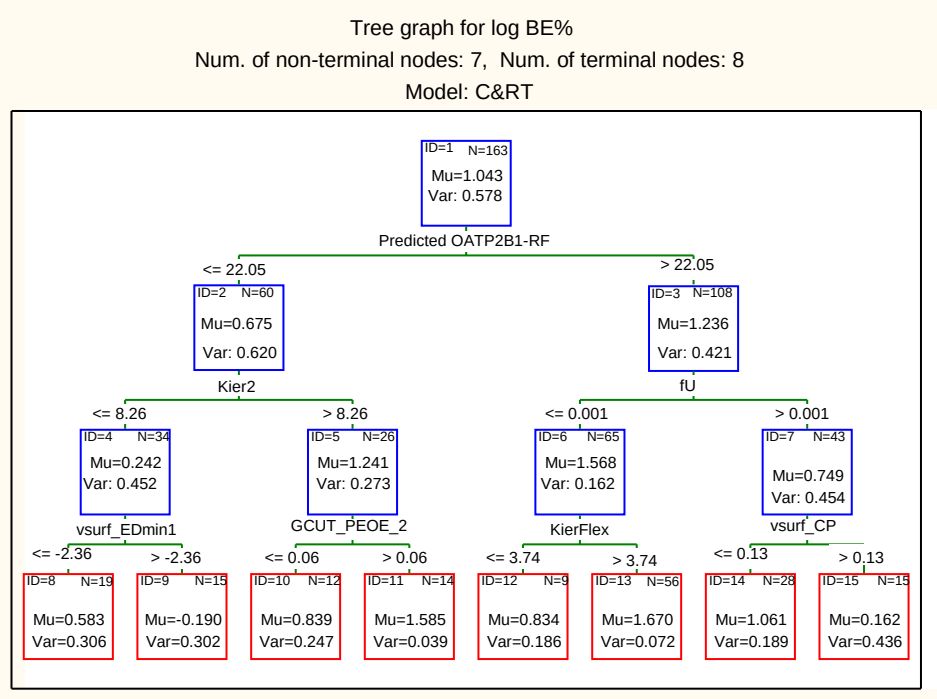

Figure 6.13. I-Tree (6) developed using interactive C&RT analysis using OATP2B1 descriptor as the first descriptor

To examine the impact of different OATP subtypes at one single model, predicted OATP1B1 and OATP1B3 effects (OATP1B1-RF and OATP1B3-BT) were imposed at the first and the second levels of a regression tree using interactive tree analysis module in STATISTICA. The best model (most accurate in the prediction of external validation set) from this exercise has been presented in Figure 6.14 (I-Tree (7)). According to this model, compounds with inhibitory effects on both OATP1B1 and OATP1B3 (14 compounds in node 7) have slightly higher biliary excretion than compounds with inhibitory effect on just OATP1B1 (compare nodes 7 and 6). Interestingly, compounds with no binding to either one of the OATPs (compounds in node 4), may still be highly excreted through bile if they have a high H-bond donor capacity (vsurf_HB4 > 150.18).

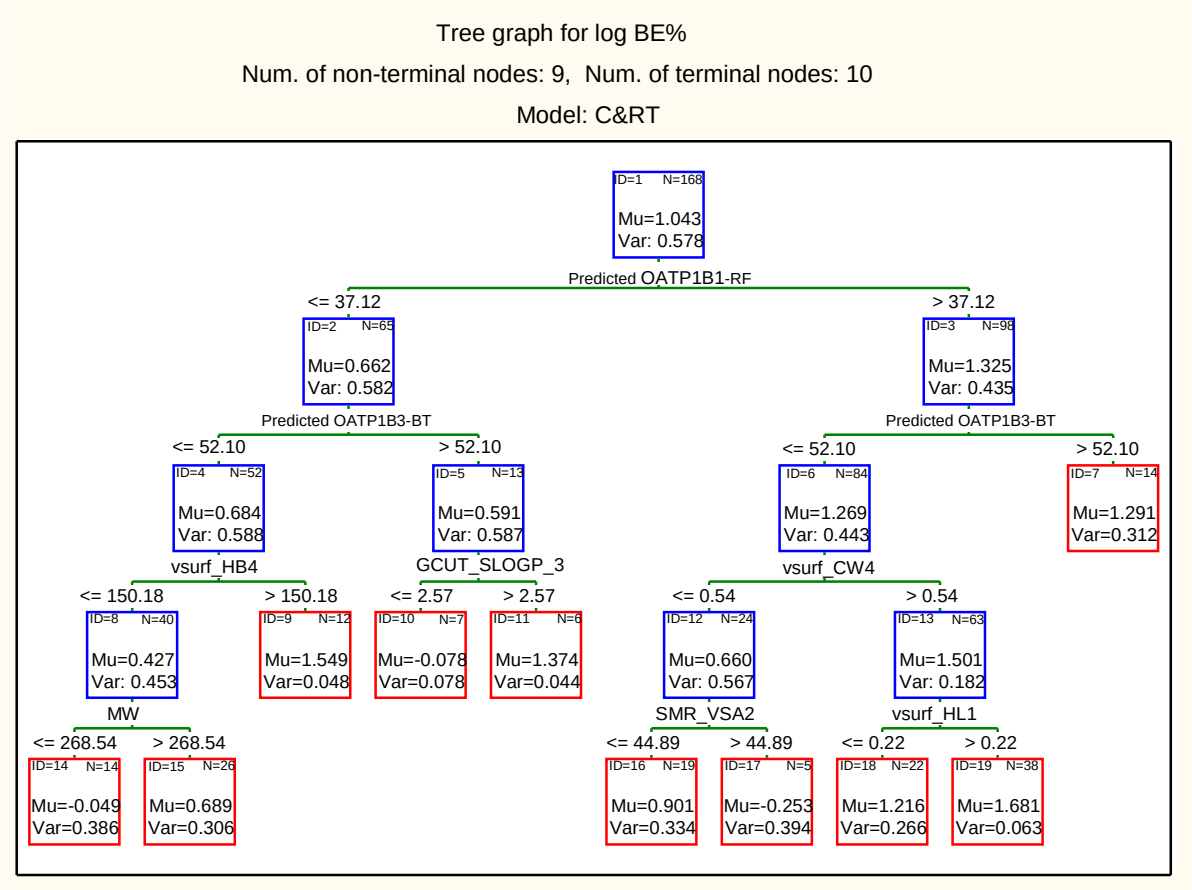

Figure 6.14. I-Tree (7) using predicted percentage OATP2B1 and OATP1B3 inhibition as the first and second level parameters

Interactive Tree Using Predicted Class

We also employed various OATP "predicted class" in the interactive tree as an alternative approach to "predicted percentage OATP inhibition" for the prediction of biliary excretion. Prediction of OATP inhibitor/non-inhibitor class for compounds in biliary excretion dataset was obtained from CT (1)-CT (3). In this way, both training and validation set compounds were predicted as class one or zero (one for inhibitor or zero for non-inhibitor). The interactive trees using predicted OATP1B1, OATP1B3 or OATP2B1 class as the first partitioning variable (I-Tree (8) – I-Tree (10)) are presented in Figures 6.15-6.17 respectively. The molecular descriptors employed in the trees have been explained in Table 6.1. Statistical parameters of these tree models can be seen in Tables 6.7 and 6.8.

I-Tree (8) in Figure 6.15 shows a slightly higher average biliary excretion for non-inhibitors of OATP1B1, which is contrary to the expectations and also different

from the result seen in I-Tree (4) employing percentage inhibition of OATP1B1 using RF (Figure 6.11). This may be due to poor prediction accuracy of CT (1) for the compounds in the biliary excretion dataset, or due to the threshold of 50% inhibition used for the classification of inhibitors/ non-inhibitors. It can be noted in I-Tree (4) that a threshold value of 35.80% (rather than 50%) has been selected by the analysis to split the compounds. Figure 6.15 also shows that both classes (inhibitors and non-inhibitors) may be divided into compounds with similarly high (nodes 5 and 7) and similarly low (nodes 4 and 6) biliary excretion using specific molecular descriptors. According to this model, in agreement to the results seen in Chapters 4 and 5 (e.g. RT (1)), compounds with large hydrophilic volume (vsurf_W3 > 418) and large hydrophilic surface ratio (vsurf_CW4 > 0.69) are excreted more in the bile.

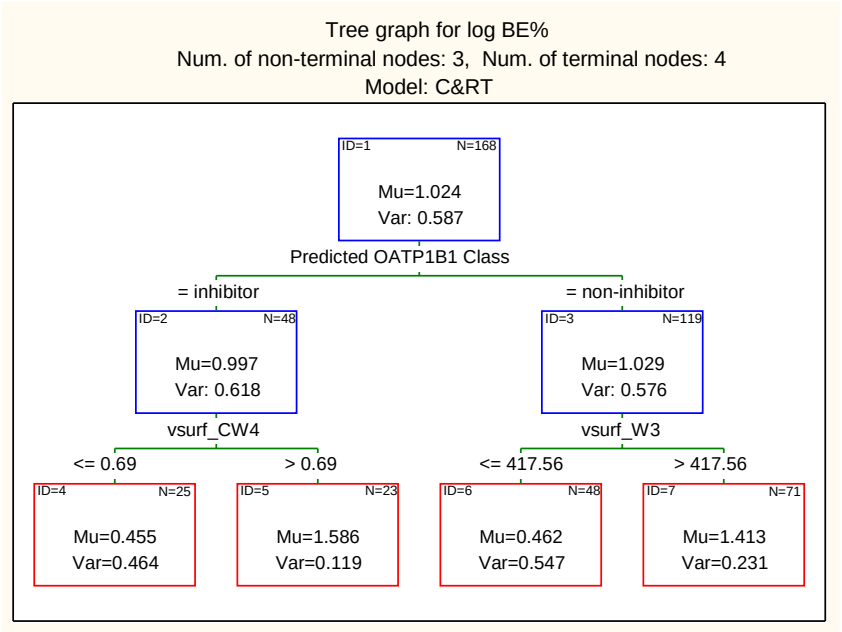

Figure 6.15. I-Tree (8) using predicted OATP1B1 inhibition class as the first parameter

Figure 6.16 (I-Tree (9)) shows that the predicted OATP1B3 inhibitor class (node 3) has higher biliary excretion, except for the compounds with extremely weak acid or base dissociations which are also composed of mainly lipophilic parts (vsurf_CP > 0.10). It can be seen in Figure 6.16, that the numbers of non-inhibitor compounds is more than inhibitors (as predicted by CT (2)) (102 vs 65). Non-inhibitors of OATP1B3 have considerable biliary excretion (terminal nodes 15, 16 and 17),

when they have high H-bond donor capacity (vsurf_HB3 > 295) and more spherical shape (glob > 0.10), or if they are not spherical, they have a strongly acidic group (Hmin > 0.64).

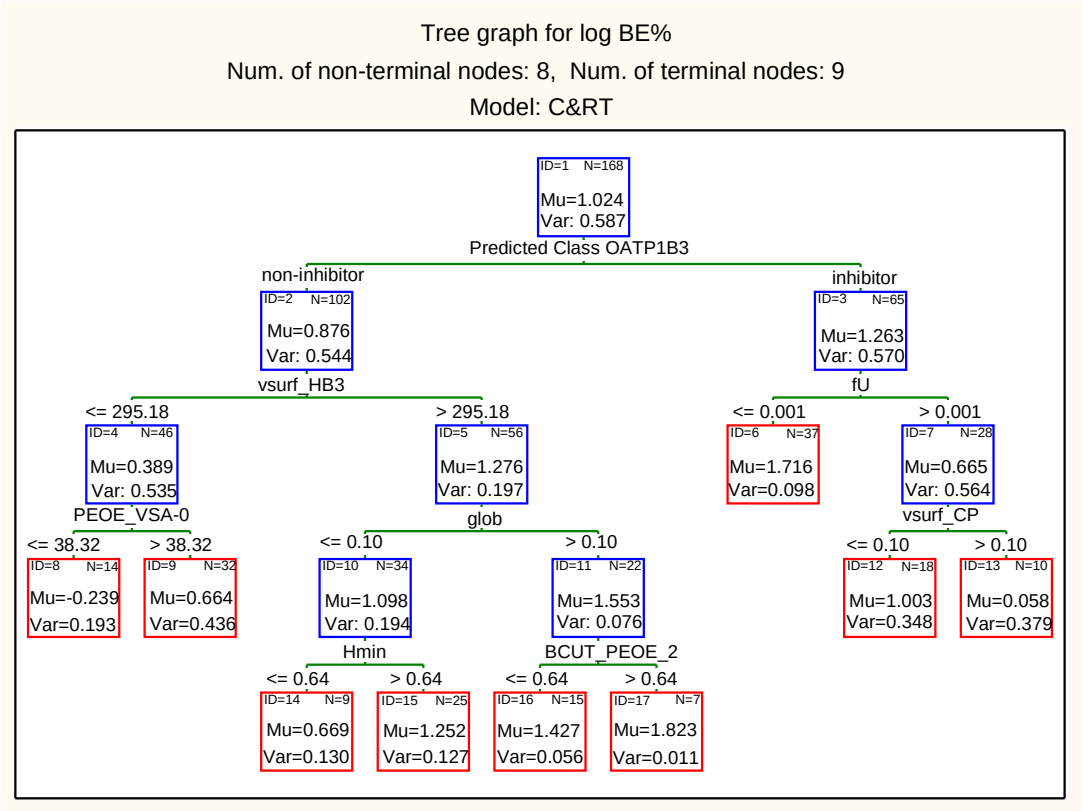

Figure 6.16. I-Tree (9) using predicted OATP1B3 inhibition class as the first parameter

I-Tree (10) in Figure 6.17 shows the effect of using predicted OATP2B1 inhibition class (by CT (3)) as the first parameter of the regression tree. According to I-Tree (10), OATP2B1 inhibitors have higher biliary excretion especially if they have a high polar surface area (PEOE_VSA+4 > 19.7). On the other hand, the 55 non-inhibitor compounds in node 4 with a low H-bond donor capacity have low biliary excretion.

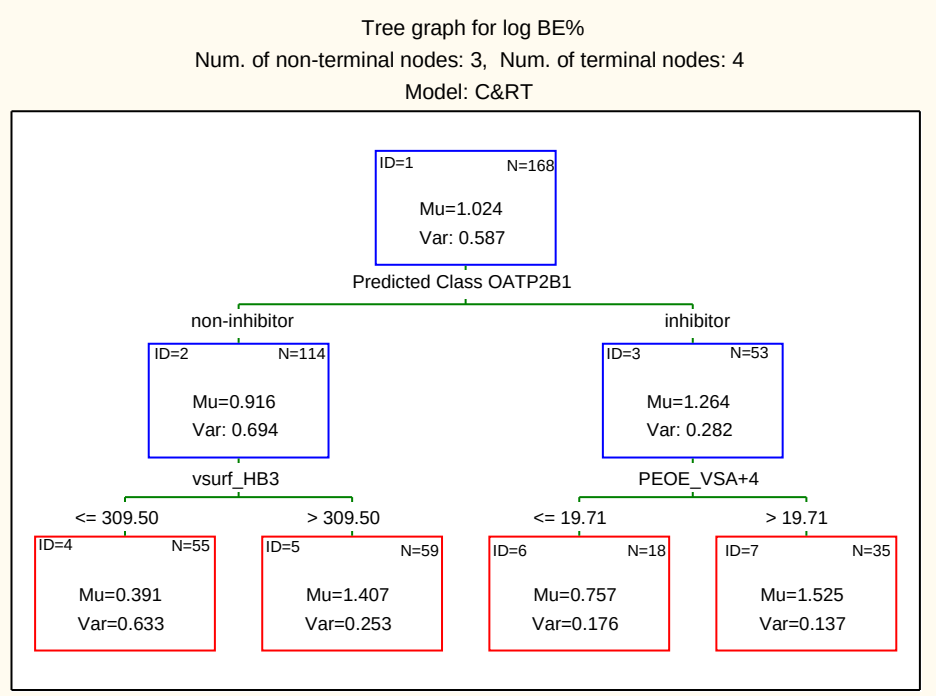

Figure 6.17. I-Tree (10) using predicted OATP2B1 inhibition class as the first parameter

6.3.3.3. Boosted Trees Model Using Predicted OATP Effects

BT analysis with various parameters as explained in Chapter 4, including various learning rates and subsample proportions were examined and the best model was selected based on the internal validation set error. The selected model (BT (5)) was obtained with the optimal number of trees of 141, learning rate of 0.1 and subsample proportion of 0.50 (see Figure 6.18).

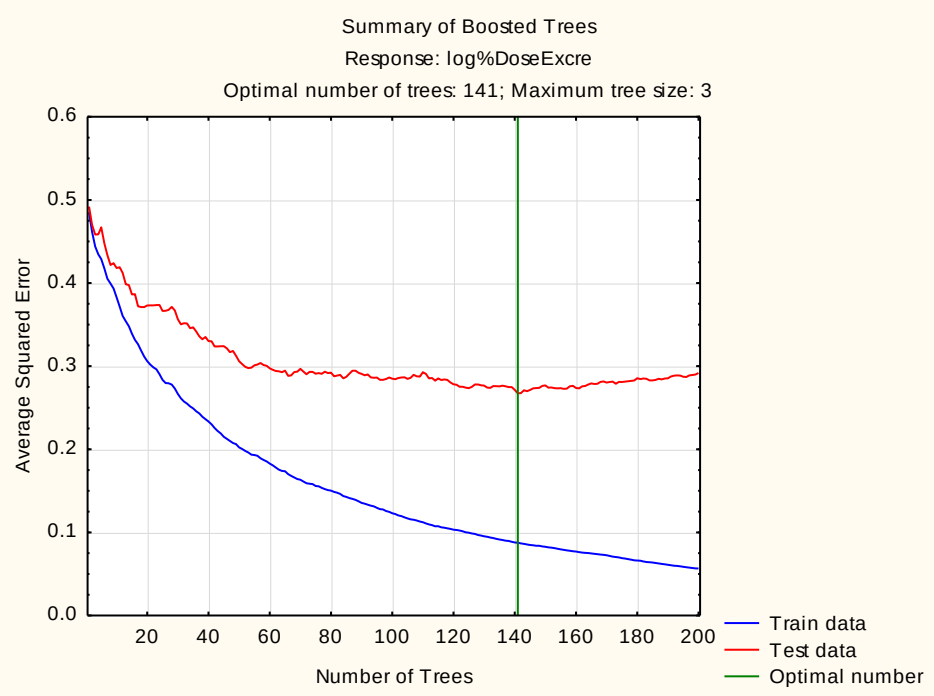

Figure 6.18. Average squared error of log BE% against the number of trees in the boosted trees model BT (5) for the training and internal test set

Variable importance was calculated for the BT model using STATISTICA software. Included in Table 6.5 are the top 10 most important molecular descriptors of BT (5) model. Lipophilicity descriptors (LogD(5.5), LogD(6.5), LogD(7.4) and LogD(10)), hydrogen atom level E-state descriptors (Hmax and Hmaxpos) and vsurf and density descriptors (vsurf_CW2 and dens) are among the top important BT (5) descriptors. Although the predicted OATP binding parameters are not amongst the top 10 descriptors of the model, they appear to be very important in this model in terms of improving the prediction accuracy for the external validation set (Tables 6.3). The previous BT models obtained from molecular descriptors (BT (1) and BT (2) in Chapter 4), and the BT model using predicted P-gp binding in addition to molecular descriptors (BT (4) in Chapter 5) have similar MAE values of 0.412, 0.417 and 0.416, respectively. BT (5) appears to be considerably more accurate with MAE of 0.362.

6.3.3.4. Random Forest Model Using Predicted OATP Effects

The method for the development of a random forest (RF) model has been explained in Chapter 4. Based on the accuracy for the internal test set, the selected RF model (RF(3)) was obtained using a subsample proportion of 0.50, numbers of trees of 100, random test data proportion of 0.2 the software's default settings for stopping conditions including minimum number of cases, maximum number of levels, minimum number in child node and the maximum number of nodes of 6, 10, 5 and 100, respectively. Figure 6.19 shows the plot of prediction error against the number of trees. Tables 6.2 and 6.3 show the statistical significance of this model.

Similar to BT model, the variables importance was calculated for RF (3). Included in Table 6.1 are the top 10 most important molecular descriptors of model. These are vsurf descriptors (vsurf_W4 and vsurf_HB1), number of single bonds (b_single and b_1rotN), kappa shape indexes (KierA2 and chi1v), number of atoms (a_count) and water accessible surface area of atoms with a negative partial charge (ASA-). Despite the absence of predicted OATP binding parameters in the top ten important parameters list, the use of these parameters in model development has resulted in a reduction in external validation set error when comparing RF (1) with MAE of 0.496 with RF (3) with MAE of 0.411.

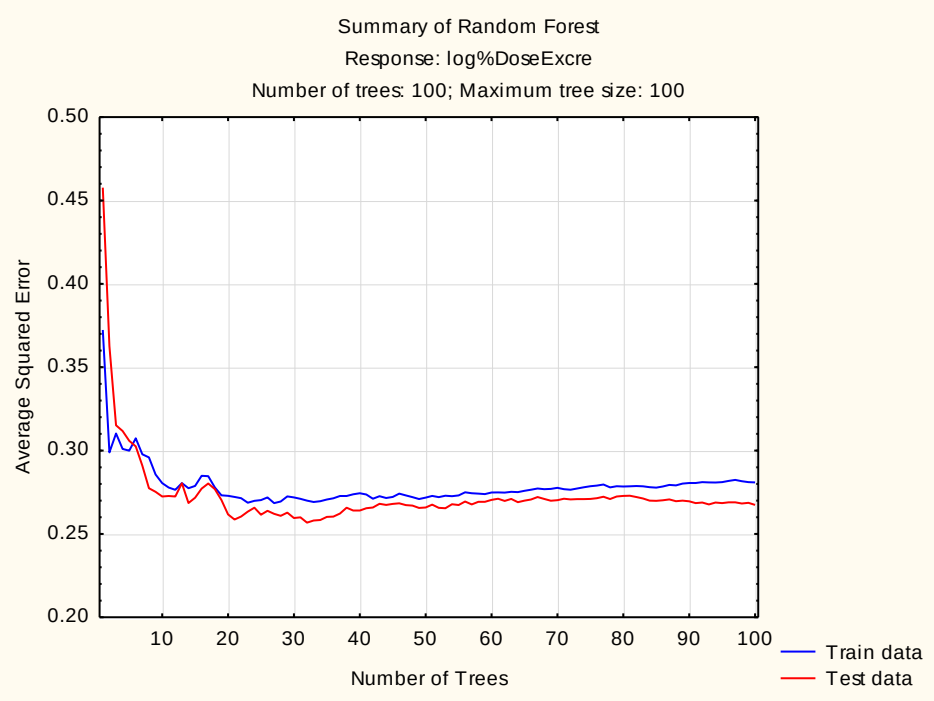

Figure 6.19. Average squared error of log BE% against the number of trees in RF (3) for the training and internal test set

6.4. Discussion

Although in the past decade the knowledge of OATP transporters had an enormous increase in the literature, most of OATP sub-family are still anonymous (Giacomini *et al.*, 2010). Various member of OATP transporter family contribute to drug disposition and, as a result, are involved in drug-drug interactions. A major contribution of OATP transporters to drug disposition is through their function in hepatocytes for the uptake of substrate compounds from the blood (Fenner *et al.*, 2012). Recently, OATP1B1 inhibition measures have been suggested as a suitable surrogate for the more complicated human hepatic uptake assays (Soars *et al.*, 2012). This was based on a comparison between uptake measures in human hepatocytes (*in vitro* intrinsic clearance) and IC_{50} values for the inhibition of OATP1B1-mediated uptake of a model substrate for 42 compounds from several chemically distinct series. In this investigation the aim was to use the OATP inhibition measured *in vitro* for the prediction of biliary excretion in rats.

6.4.1. QSAR Models for the Prediction of OATP Inhibition

Despite the wide distribution and important implications of OATP transporter family, unfortunately, there are several limitations in the study of OATP transporter ligands (Karlgren *et al.*, 2012a). This has resulted in a limitation in the availability of high quality data for QSAR studies. In this investigation, the inhibition of OATP uptake of a substrate by 225 compounds measured as percentage inhibition by a single concentration of the compound (Karlgren *et al.*, 2012a) was used as the inhibition measure. Data was available for three major OATP subfamilies, OATP1B1, OATP1B3 and OATP2B1. OATP1B1 and OATP1B3 are liver-specific transporters, mainly expressed on the basolateral membrane of human hepatocytes (Kalliokoski and Niemi, 2009; Giacomini *et al.*, 2010), whereas, OATP2B1 is relatively ubiquitous with its localization in several tissues in addition to the liver (Kobayashi *et al.*, 2014; Varma *et al.*, 2011).

After examining several prediction (regression based) statistical techniques (stepwise regression analysis, C&RT, BT, RF and MARS), the two best models were selected for each OATP subfamily. In addition a classification tree was developed for each subfamily, using 50% inhibition as the threshold value for inhibitors/ non-inhibitors.

OATP1B1 Inhibitors

For OATP1B1, RF and C&RT analysis resulted in the best prediction models (OATP1B1-RF and OATP1-RT). There is only one molecular descriptor used in OATP1-RT model, Chi1_C, which is mainly an indicator of molecular size. Despite previous investigations suggesting that ligands of this transporter are mainly acidic (Hsiang *et al*., 1999) this has not been indicated in this model. In comparison with the regression tree, the classification model for OATP1B1 (CT (1)) has more branches and nine terminal nodes. The importance of acidic nature of OATP1B1 ligands has been indicated in CT (1). In CT (1), in order to be classed as inhibitors, compounds of smaller size (defined by Chi1_C < 9.68) need to have acidic group shown by partially positively charged hydrogen, as in –COOH group (Hmin), or high apparent partition coefficient in acidic pH (logD(2)). The crucial impact of

large molecular size for OATP ligands is very well established from previous studies. Whereas OATs transport low MW compounds, OATPs mediate the uptake of larger substrates such as digoxin (Shitara *et al.*, 2002; Hagenbuch and Meier, 2003), erythromycin (Sun *et al.*, 2004) and atorvastatin (Lau *et al.*, 2006). This is also in line with a study by Hagenbuch and Meier which reports that compounds with molecular weight higher than 350 can be OATP1B1 substrates (Hagenbuch and Meier, 2004).

A recent QSAR model by Soars and colleagues using IC_{50} values for 262 proprietary compounds found that maximal hydrogen bonding strength and lipophilicity (cLogP) were the most important molecular descriptors of their random forest model for predicting OATP1B1 inhibitors (Soars *et al.*, 2014). Our random forest model also supports this finding as lipophilicity (LogP) and maximum positive hydrogen atom-level E-state value in a molecule (Hmaxpos) were dominant molecular features in OATP1B1-RF model. In addition, CT (1) also suggests the importance of lipophilicity (LogD(2)) for inhibitors of OATP1B1. De Bruyn and co-workers in a recent study, noted the polar surface area as the key molecular feature for an increase in OATP1B1inhibition (De Bruyn *et al.*, 2013), which is in agreement with CT (1) indicating the positive impact of a high hydrophilic/lipophilic balance of the molecular surface (vsurf_HL1) and a large negative polar surface area (PEOE_VSA_NEG) for the compounds to be classed as inhibitors of OATP1B1.

The accuracy of the regression based models for the external validation set is similar to the training set (MAE for the percentage inhibition is ~21%). This percentage error must be viewed considering the innate error levels associated with the single point measurements. Karlgren et al. (2012) have developed classification, rather than regression based, QSAR models using this dataset. Their classification accuracy for the training and validation sets was 73% and 79% respectively, which is similar to CT (1) model (accuracy of 81% for inhibitors and 74% for non-inhibitors in the external validation set).

OATP1B3 Inhibitors

The selected regression based models for OATP1B3 inhibition were a RF and a BT model (OATP1B3-RF and OATP1B3-BT). Despite allowing for identification of the most important features, these two methods cannot be interpreted as directly as single classification or regression trees. CT (2) has a very low classification accuracy for the non-inhibitors in the external validation set (36%), despite performing well for the classification of inhibitors in the same set (83%). Therefore, consideration must be given to the accuracy levels when interpreting the molecular properties of inhibitors and non-inhibitors. An inspection of CT (2) provides required features for inhibitors as explained in Section 6.3.2. Mainly, the inhibitors are either flexible with a relatively small fraction of polar surface area, or they are more rigid with large negative polar surface area or with a specific molecular topology with various BCUT descriptors. The BCUT descriptors have been reported to be very useful in terms of capturing sufficient structural detail in molecular diversity-related tasks (Stanton, 1999; Pearlman and Smith, 1997). Despite this, the incorporation of this parameter to explain variations in the biological properties is not successful in this model.

As explained in the resuts section, the most important molecular descriptors of OATP1B3-BT are LogD at pH 10, acidity, aromatic rings, and hydrophilicity or hydrogen bonding descriptors. This is in agreement with the findings of De Bruyn and co-workers that indicate a LogD value between 3.4 and 7.5 and a medium/ low number of hydrogen bond donors are positively correlated with OATP1B3 activity (De Bruyn *et al.*, 2013). The most important molecular descriptors of OATP1B3-RF are similar to CT (2) model and indicate the importance of the bond count and the number of single bonds. In addition, this model also indicates the importance of hydrogen bonding donor capacity, molecular shape, and volume. The prediction accuracy of the regression based OATP1B3 models is similar to the models for OATP1B1 at ~20% for the external validation set.

OATP2B1 Inhibitors

A recent study by Shirasaka and colleges (Shirasaka *et al.,* 2014) on OATP2B1-mediated uptake of pravastatin and fexofenadine showed the presence of multiple binding sites on OATP2B1. The structure of OATP2B1 has been shown to be very similar to OATP1B3 using in silico homology modeling studies (Meier-abt *et al.,* 2005), which suggest that most OATPs share similar features. Very few literature data are available for OATP2B1 ligands. For instance, out of 45 OATP2B1 inhibitors identified in Karlgren's investigation, 29 compounds were were believed to be novel inhibitors not studied before (Karlgren *et al.,* 2012a). As a result, despite a few QSAR/ pharmacophore models published for OATP1B1 (Chang *et al.*, 2005; De Bruyn *et al.*, 2013; Soars *et al.*, 2012; Karlgren *et al.*, 2012b), there is little in silico results available for OATP2B1 (Karlgren et al., 2012a). Based on the similarities with other OATP transporters, it may be speculated that OATP2B1 pharmacophores may share the similar molecular features for the consideration of the substrate binding at the positively-charged region (El-Kattan and Varma, 2012). Its substrates may have features such as a hydrophobic core to form the π-stacking interaction with the imidazole ring of amino acid H579, or a hydrogen bond donor group to directly interact with the nitrogen atom of the imidazole ring (El-Kattan and Varma, 2012).

The selected regression based models for OATP2B1 ligands are RF and BT models (OATP2B1-RF and OATP2B1-BT) and CT (3) is the classification model. CT (3) model has correctly classified 77% and 58% of the inhibitors and non-inhibitors in the external validation set, respectively. The accuracy of the PLS-based classification model suggested by Karlgren et al. (2012a) for this transporter was 75%, but they had used a different classification cut-off point of 32%. CT (2) model indicates that inhibitors of OATP2B1 are generally large hydrophilic molecules or otherwise they have a specific topological property defined by a GCUT molecular descriptor.

Both regression based models for OATP2B1 had a prediction error of ~25% (MAE = 25 for percentage inhibition data) for the external validation set (see Table 6.3). It can be seen in the results section that both these models show the importance of hydrogen bond donor ability with the molecular descriptors Hmaxpos and Hmax.

Moreover the importance of polarity is shown with polar surface area and negative polar surface area, and ratio of carbon atoms.

In brief, physicochemical variables detected as important for inhibition of each OATP sub-family, show similarities but there are also some differences observed.

6.4.2. Effect of OATP Binding on Biliary Excretion Models

For hepatobiliary elimination of compounds, it has now become progressively clear that the movement of solutes and compounds into and out of cells is often dependent on transporter proteins. After compounds enter the hepatocytes, they either undergo the metabolism process, or, the intact compounds or their metabolite molecules excrete into the bile canaliculus. The uptake transporters enhance biliary excretion by importing more compounds into hepatocytes. Among the various uptake transporters, OATP family members appear to have remarkably broad substrate specifications (Kim, 2003). In human and rat hepatocyte, the hepatic uptake of many compounds is mediated by OATP family. Nevertheless, the physiological role of the OATP family is still not fully understood (Mikkaichi *et al.*, 2004). Varma et al (2012) in their research paper comparing biliary excretion of compounds and the chemical space of substrates of human OATPs and rat oatp1b2 observed that there is a significant overlap between these substrates and compounds with a rat biliary excretion higher than 10%.

In this investigation, the predicted OATP inhibition values were used as parameters (predictors) for the development of QSAR models for the biliary excretion of compounds. In assessing the effect of predicted OATP binding on the QSAR models for biliary excretion, it must be noted that QSAR has been used for the prediction of OATP effect and that these original OATP QSARs are based on percentage inhibition data which is a fast measure of inhibition activity but is less reliable than IC_{50} values.

Using C&RT embedded feature selection, only OATP1B3 inhibition is selected in the tree structure, and even this is at lower branches of the tree indicating less significance of the parameter (RT (3) in Figure 6.10). Moreover, the effect seen by

this parameter is in contrast to the expectations that a higher OATP binding should result in higher biliary excretion. It must be noted here that the number of OATP binding parameters (two numerical predicted percentage inhibition and one categorical inhibition class for each subfamily of OATP, making nine in total) is much lower than the number of molecular descriptors used (more than 300 in total). This gives a higher statistical probability to the molecular descriptors to be selected by any statistical feature selection. The OATP descriptors were therefore incorporated in the tree structure manually using Interactive Tree analysis in STATISTICA. Table 6.8 gives the details of I-Tree (4) – (10) models, and Table 6.7 gives the prediction accuracy for the training and external validation sets. Table 6.7 shows that I-Tree models (8) – (10), using the categorical predicted class variables are less accurate than the corresponding I-Tree (4) – (7) using the numerical predicted percentage OATP inhibition. This may indicate a higher prediction accuracy for the regression based models for the prediction of OATP effect of compounds in the biliary excretion dataset.

Among the OATP member family, the role of OATP1B1 in elimination of compounds has become clear over the last decade (Soars *et al.*, 2012). Accordingly, comparing accuracy of I-Tree (4) – (7), it is clear that, out of different OATP subfamilies, incorporation of OATP1B1 inhibition results in the most successful model (I-Tree (4) followed by I-Tree (7)). Moreover, incorporation of predicted OATP2B1 subfamily results in the least accurate model (I-Tree (6)). This may be due to a lower prediction accuracy of the original OATP2B1 model (OATP2B1-RF in Table 6.3 with MAE of 25%) rather than a lower significance of OATP2B1 binding in hepatic uptake and biliary excretion.

It can be seen that the prediction accuracy of I-Tree (4) is better than RT (3) with statistically selected variables. I-Tree (4), indicates that, in general, OATP1B1 ligands have higher biliary excretion and, in addition to this, eight different levels of log BE% values may be identified by this tree based on several molecular properties. The molecular properties have been explained in the results (section 6.3.3.2) and are similar to the observations from Chapter 4.

The best QSAR model for the estimation of biliary excretion, using the predicted OATP binding in addition to the molecular descriptors as the predictors, was

achieved by the boosted trees model, BT (5). BT (5), with incorporation of predicted OATP binding effects along with molecular descriptors, is much more accurate than the corresponding BT (1) and BT (2), with only molecular descriptors, and BT (4), with incorporation of P-gp binding and molecular descriptors.

Since the biliary excretion dataset is completely external and there is no OATP data for these compounds, it is difficult to comment on the prediction accuracy of OATP inhibition for this dataset using the QSAR models other than the error indication given by the external validation set (MAEs reported in Table 6.3 and SP and SE values in Table 6.4). In terms of the chemical space, there seems to be a good overlap between the molecular properties of the two training sets, as indicated by a visual inspection of the scores plot from principle component analysis (PC1 vs PC2 plot in Figure 6.20).

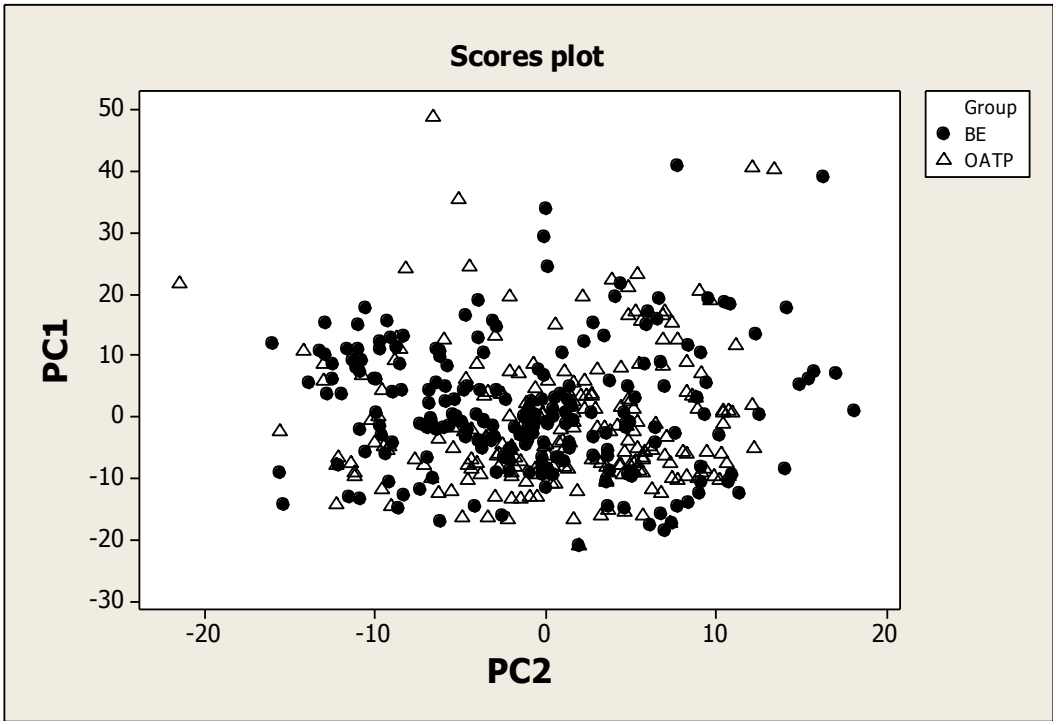

Figure 6.20. The plot between the first and the second principle components of PCA using all the molecular descriptors

In conclusion, incorporation of OATP effects in the prediction of biliary excretion resulted in better regression tree models when incorporated manually in interactive

trees. Furthermore, a BT model was achieved when OATP effects were used in addition to molecular descriptors as predictors of biliary excretion.

7. General Conclusion

Biliary excretion is one of the major elimination routes for drugs and as a result, it has a major impact on pharmacokinetics including drug half-life and dosing regimen. Moreover, biliary excretion has implications in drug-drug and food-drug interactions through the possible involvement of same transporter proteins. As a result, early estimation of biliary excretion may be useful for modification of drug structure in drug design to have an ideal drug and can be used as a surrogate for more time-consuming and expensive *in vivo* and *in vitro* studies. In this project, we were able to estimate rat biliary excretion based on physicochemical properties using various computational modelling techniques. In addition, the roles of P-gp and OATPs, as two important hepatobiliary influx and efflux transporters were investigated using QSAR.

The statistical techniques used for the QSAR development included a range of linear, non-linear and ensemble methods to allow the best possible prediction accuracy. The methods were multiple linear regression analysis, decision trees developed by C&RT and CHAID, MARS, and ensemble decision trees developed by random forest and boosted trees methods. Simple models such as classification or regression trees, multiple regression analysis and MARS, use manageable number of features and allow for easy interpretation of the results. In this way, the selected molecular descriptors resulted in some insight into major factors that can affect biliary elimination of drugs.

The biliary excretion dataset used in this project consisted of a diverse dataset of 217 compounds with percentage of dose excreted intact into bile measured *in vivo* in rat. The first aim of the investigation was to develop a predictive QSAR model for this dataset. Table 7.1 gives a brief summary of the prediction accuracy of all the biliary excretion models described in this thesis. The most accurate models in terms of the prediction accuracy for the external validation set in descending order of accuracy are CHAID (2), BT (5), RT (1) and I-Tree (4). This shows that simple regression trees such as CHAID (2) and RT (1) are as powerful in the prediction of

biliary excretion as the more sophisticated ensemble methods of boosted trees and random forest techniques.

Table 7.1 MAE values of all the biliary excretion models described in the thesis; the selected models have been highlighted in bold.

Model	Training set	Validation set
BT (1)	0.229	0.412
BT (2)	0.226	0.417
BT (4)	0.339	0.416
BT (5)	**0.242**	**0.362**
CHAID (2)	**0.432**	**0.359**
I-tree (1)	0.345	0.451
I-Tree (10)	0.448	0.474
I-tree (2)	0.424	0.468
I-Tree (4)	**0.343**	**0.379**
I-Tree (5)	0.335	0.409
I-Tree (6)	0.332	0.443
I-Tree (7)	0.362	0.392
I-Tree (8)	0.454	0.455
I-Tree (9)	0.334	0.446
MARS (2)	0.438	0.428
MARS (3)	0.436	0.442
MLR (1)	0.377	0.483
RF (1)	0.403	0.496
RF (3)	0.387	0.411
RT (1)	**0.304**	**0.373**
RT (3)	0.236	0.420

From these models, we obtained an insight into the structural profile of cholephilic compounds through accurate modelling of the biliary excretion. Molecular descriptors selected by all these models including the top ten incorporated in boosted trees and random forest models indicated a higher biliary excretion for relatively hydrophilic compounds especially if they have acid/base dissociation (anionic or cationic), and have a large molecular size.

Interactive regression trees analysis was a very useful tool that helped investigate the effects of specific properties. One such property with regards the previous

literature was molecular weight. Despite the established role of molecular weight in biliary excretion, the molecular weight thresholds in previous literature are generally based on qualitative inference from available data, rather than a statistically established threshold (Yang *et al.*, 2009). In this project a statistically validated molecular weight threshold established for significant biliary excretion at MW = 348 Da.

Analysis of outliers in majority of the models in Chapter 4 showed the models perform best when lipophilicity is not too extreme (log P < 5.35) and for compounds with molecular weight above 280 Da. It was also observed that compounds with low biliary excretion are more likely to show a higher average error. This could be attributed at least in part to the method used for calculation of error as, for example, despite the prediction of low biliary excretion at 1% for a compound, the difference with the observed value of 0.1% leads to a high absolute error of 1. Such estimations may still be acceptable as these low biliary excretion compounds had been estimated a BE% value < 4%.

P-gp is a major efflux pump that operates in hepatocytes and aids with excretion of its substrates into bile. Based on the hypothesis that the substrates of this transporter may have a higher tendency to be excreted through bile, this project looked at the structural features of P-gp ligands. A very accurate measure of ligand binding to proteins is the inhibition constant (K_i). K_i is believed to be a more universal parameter allowing easy comparison of data from different substrate conditions. To investigate the molecular requirements of P-gp binding and the effect of P-gp binding on biliary excretion levels of compounds, a dataset of 219 unique P-gp inhibitor/substrate pairs were collated from original literature. QSAR models were developed for K_i using P-gp-ligand docking scores as well as the molecular descriptors of the inhibitors and the descriptors of probe substrates used for the determination of K_i values. The QSARs indicated that the molecular descriptors are more significant in the prediction of P-gp binding than the ligand-enzyme docking scores. Models indicated that the potent P-gp inhibitors have higher lipophilicity and molecular weights than drug-like molecules identified by Oprea's rule. The QSAR models indicate that potent inhibitors of P-gp have higher lipophilicity and molecular size than lead-like compounds as defined by Oprea and the limiting

lipophilicity is log P > 5.3 for this dataset. Classification and regression tree (C&RT) model had the lowest K_i prediction error for the external validation set with a mean absolute error of 0.543.

Although the QSARs established for P-gp had reasonable accuracy for the prediction of K_i values of the external validation set, these predictions may not be as reliable for the external compounds in the biliary excretion dataset. This can occur in case the compounds in the biliary excretion dataset are outside the domain of applicability of the QSAR models for P-gp binding. A scores plot from PCA showed a considerable difference between the chemical spaces of the two datasets. Therefore it was not unexpected when the predicted P-gp inhibition constant could not significantly improve the prediction accuracy of the QSAR models.

OATPs are major uptake transporters that mediate the uptake of a wide range of compounds from blood into hepatocytes as the first step of hepatobiliary elimination process. To study the significance of OATP binding in biliary excretion, a recently published dataset consisting of percentage inhibition of three OATP subtypes, OATP1B1, OATP1B3 and OATP2B1 by 225 compounds was employed. Despite the lower quality of this binding measure in comparison with IC_{50} or K_i, QSARs of reasonable accuracy (MAE of 20-25%) were established for the three OATP subtypes. In addition, a classification method, i.e. classification tree, was also used. Both regression type and classification methods were most successful for the prediction of OATP1B1 binding when compared to OATP1B3 and OATP2B1 binding. This may be attributed to a more balanced inhibitor/non-inhibitor ratio in the dataset for this particular OATP. The results showed large hydrophilic compounds with hydrogen bonding donor ability (such as carboxylic acid groups) are better inhibitors of OATP1B1 and OATP2B1, while flexibility was an additional factor for OATP1B3.

A comparison of the chemical spaces of compounds in OATP dataset with compounds in biliary excretion dataset using PCA indicated a good overlap of properties. The OATP models were used for the prediction of OATP binding of the compounds in biliary excretion dataset and the predicted values were used as additional parameters for the estimation of biliary excretion using QSAR. Although

majority of these predicted OATP binding parameters were not picked by C&RT algorithm, and they were not ranked within the top ten most important features of BT or RF models, they were important in improving the prediction accuracy of BT model and the regression trees, when they were incorporated manually using interactive trees. In the selected I-Tree model, the predicted OATP1B1 binding was the most significant parameter and this constitutes one of the best models over all for the prediction of biliary excretion with an absolute error of 0.38 (I-Tree (4), Table 7.1). The BT model has a slightly lower prediction error of 0.36 for the external validation set.

8. Future Work

As a result of the research carried out in this PhD project, it can be seen that there is a need to further explore the role of individual ABC transporters as the efflux pumps. In addition to the role and impact of efflux drug transporters in the hepatocyte, further investigation of the impact of both uptake and other efflux hepatic transporters in biliary excretion as well as search for new transporter dataset for biliary excretion such as Pept1, Pept2, MRP2, MRP3, MRP4, MRP6, MATE1, OAT2, OAT7, OCT1, NTCP, BSEP, PHT1 and PHT2 can elucidate and bring more clear aspects of elimination pathways to light.

In terms of P-glycoprotein there are large datasets of substrate/non-substrate type, some of which are proprietary data and some (smaller datasets) are available in the literature (Wang *et al.*, 2011; Broccatelli *et al.*, 2011). Although the data is categorical which is not ideal, the chemical space of these datasets may be closer to the compounds in biliary excretion dataset. In addition to P-gp, two other efflux pumps are also very important in biliary excretion. These are MRP2 and BCRP which have high localisation in hepatocytes. The work may involve cutting edge QSAR models along with classic QSAR model development, as well as drug-enzyme docking methods. These transporter enzymes have also been indicated to play roles in the anticancer drug resistance and also pharmacokinetic processes such as intestinal absorption and blood brain barrier transport. Therefore the models will be useful from other perspectives as well as biliary excretion.

The lack of high resolution structures of several important transporters including P-glycoprotein and OATPs has severely limited work in this field. For example, if higher resolution models of P-glycoprotein were made available, this may improve the docking energies and allow us to visualise the interactions between P-glycoprotein and compounds. In terms of P-gp docking, in this work the binding pocket was defined using the location of a single co-crystallised ligand. P-gp is known to have several binding sites and can accommodate more than one ligand at a time. A more detailed investigation may look at docking at several binding sites and then, from QSAR perspective, the lowest energy binding could be selected for

each compound from the various binding sites to be used as a QSAR parameter. Besides, building structure-based pharmacophore models of P-glycoprotein especially with pharmacophore features of hydrophobic, aromatic rings, hydrogen bond acceptors or donor, cations, and anions can be helpful.

In order to further confirm the external applicability and predictive ability of the models built in this study as good predictors of P-glycoprotein and OATP binding and predictors of biliary excretion, new sets of compounds should be used as external validation set to test the constructed models. A major practice, which should be carried for the models presented in this thesis, is to investigate diversity of the compounds in the datasets and to define the applicability domain of the models.

Furthermore, it will be pertinent to ensure that datasets are robust. For example, for P-glycoprotein substrates the goodness of the methods used for the measurement of activity should be scrutinised, and several sources of data should be compared if compounds or dataset to be used for model building have been repeatedly identified in several studies as either substrates or non-substrates of P-glycoprotein.

Apart from key continuous and classification computational methods for estimation of biliary excretion used in this study, other statistical techniques can be utilized to predict the biliary excretion e.g. neural networks, support vector machines and semi-supervised learning. Neural networks and support vector machines can be used as a helpful alternative when there are problems of prediction or classification. Semi-supervised learning is a class of supervised learning techniques that make use of unlabelled data for training and has emerged as an exciting new direction in machine learning research. For example, in the biliary excretion dataset, when biliary excretion values are converted to log BE%, there are a few missing values for a few compounds (nine) with zero biliary excretion. Semi-supervised learning methods can improve models generalizability and applicability by predicting the values for these compounds.

In this investigation, we searched for biliary excretion or clearance data for other species before analysis of rat biliary excretion database. For human, we could collect a biliary excretion data of 68 compounds. There are some biliary excretion

data available for dog and rabbit. However, we did not analyse these datasets owing to the limited number of compounds in the datasets. As a part of future work of this thesis, to cope with the lack of human biliary excretion dataset, we suggest the extrapolation to human pharmacokinetic parameters mainly from rat data (but also from dog, and monkey data).

It should be noted that the uptake of drugs via the sinusoidal membrane and drug efflux by transporters is a complicated process; further studies of transporter-mediated drug-drug interaction in hepatocyte, additional investigation on *in silico* and *in vitro* transporter methods, linking and utilising the pharmacokinetic parameters which will affect the net hepatic clearance such as area under the curve (AUC), excretion rate and ratio and half-life is necessary and can elucidate the overall elimination process in the liver hepatocyte.

The relationship between biliary excretion and hepatic metabolism is beyond the scope of the present study, however, this should be possible with more data on metabolism and using statistical techniques such as partial least squared regression (PLS) which allows predicting more than one variable at the same time.

Finally, the biliary excretion, OATPs, K_i, K_m and IC_{50} dataset can be populated with more data as they become available in the literature.

9. References

Abe T, Kakyo M, Tokui T, Nakagomi R, Nishio T, Nakai D, Nomura H, Unno M, Suzuki M, Naitoh T, Matsuno S, Yawo H. 1999. Identification of a novel gene family encoding human liver-specific organic anion transporter LST-1. J Biol Chem. 274(24)17159-63.

Abe T, Unno M, Onogawa T, Tokui T, Kondo TN, Nakagomi R, Adachi H, Fujiwara K, Okabe M, Suzuki T, Nunoki K, Sato E, Kakyo M, Nishio T, Sugita J *et al.* 2001. LST-2, a human liver-specific organic anion transporter, determines methotrexate sensitivity in gastrointestinal cancers. Gastroenterology. 120(7)1689-99.

Abou-El-Makarem MM, Millburn P, Smith RL, Williams RT. 1967. Biliary excretion in foreign compounds.Benzene and its derivatives in the rat. Biochem J. 105(3)1269-74.

Adjepon-Yamoah KK, Nimmo J, Prescott LF. 1974. Gross impairment of hepatic drug metabolism in a patient with chronic liver disease. Br Med J. 4(5941)87-88.

Agnani D, Acharya P, Martinez E, Tran TT, Abraham F, Tobin F, Ellens H, Bentz J. 2011. Fitting the elementary rate constants of the P-gp transporter network in the hMDR1-MDCK confluent cell monolayer using a particle swarm algorithm. PLoS One. 6(10)1-12.

Ahlin G, Karlsson J, Pedersen JM, Gustavsson L, Larsson R, Matsson P, Norinder U, Bergström CA, Artursson P. 2008. Structural requirements for drug inhibition of the liver specific human organic cation transport protein 1. J Med Chem. 51(19)5932-42.

Akamatsu M. 2002. Current state and perspectives of 3D-QSAR. Curr Top Med Chem. 2(12)1381-94.

Allen JD, Brinkhuis RF, Wijnholds J, Schinkel AH. 1999. The mouse Bcrp1/Mxr/Abcp gene: amplification and overexpression in cell lines selected for resistance to topotecan, mitoxantrone, or doxorubicin. Cancer Res. 59(17)4237-41.

Aller SG, Yu J, Ward, Weng, Chittaboina S, Zhuo, Harrell PM, Trinh YT, Zhang Q, Urnatsch IL, Chang G. 2009. Structure of P-glycoprotein reveals a molecular basis for poly-specific drug binding. Science. 323(5922)1718-22.

Alonso H, Blizniyuk AA, Gready JE. 2006. Combining Docking and Molecular dynamic simulations in drug design. Med Res Rev. 26(5)531-68.

Amacher D, Schomaker S, Retsema J. 1991. Comparison of the effects of the new azalide antibiotic, azithromycin, and erythromycin estolate on rat-liver cytochrome P-450. Antimicrob Agent Chemother. 35(6)1186-90.

Ambudkar SV, Dey S, Hrycyna CA, Ramachandra M, Pastan I, Gottesman MM. 1999. Biochemical, cellular, and pharmacological aspects of the multidrug transporter. Annu Rev Pharmacol Toxicol. 39(1)361-98.

Ambudkar SV, Kimchi-Sarfaty C, Sauna ZE, Gottesman MM. 2003. P-glycoprotein: from genomics to mechanism. Oncogene. 22(47)7468-85.

Apiwattanakul N, Sekine T, Chairoungdua A, Kanai Y, Nakajima N, Sophasan S, Endou H. 1999. Transport properties of nonsteroidal anti-inflammatory drugs by organic anion transporter 1 expressed in Xenopus laevis oocytes. Mol Pharmacol. 55(5)847-54.

Arceci RJ, Croop JM, Horwitz SB, Housman D. 1988. The gene encoding multidrug resistance is induced and expressed at high levels during pregnancy in the secretory epithelium of the uterus. Proc Natl Acad Sci USA. 85(12)4350-54.

Arimori K, Kuroki N, Hidaka M, Iwakiri T, Yamsaki K, Okumura M, Ono H, Takamura N, Kikuchi M, Nakano M. 2003. Effect of P-glycoprotein modulator, cyclosporin A, on the gastrointestinal excretion of irinotecan and its metabolite SN-38 in rats. Pharm Res. 20(6)910-7.

Asaka JI, Terada T, Tsuda M, Katsura T, Inui KI. 2007. Identification of Essential Histidine and Cysteine Residues of the H^+/Organic Cation Antiporter Multidrug and Toxin Extrusion (MATE). J Mol Pharmacol. 71(6)1487-93.

Ayrton A, Morgan P. 2001. Role of transport proteins in drug absorption, distribution and excretion. Xenobiotica. 31(8-9)469-97.

Bahn A, Hagos Y, Reuter S, Balen D, Brzica H, Krick W, Burckhardt BC, Sabolic I, Burckhardt G. 2008. Identification of a new urate and high affinity nicotinate transporter, hOAT10 (SLC22A13). J Biol Chem. 283(24)16332-41.

Bakos E, Homolya L. 2007. Portrait of multifaceted transporter, the multidrug resistance-associated protein 1 (MRP1/ABCC1). Pflugers Arch. 453(5)621-41.

Balaban AT. 1982. Highly Discriminating Distance-Based Topological Index. Chem Phys Lett. 89(5)399–404.

Bardsleey-Elliot A, Plosker GL. 2000. Nelfinavir an update on its use in HIV infection. Drugs. 59(3)581-620.

Beaulieu E, Demeule M, Ghitescu L, Beliveau R. 1997. P-glycoprotein is strongly expressed in the luminal membranes of the endothelium of blood vessels in the brain. Biochem J. 326(2)539-44.

Benet LZ, Broccatelli F, Oprea TI. 2011. BDDCS applied to over 900 drugs. AAPS J. 13(4)519-47.

Bentz J, O'Connor MP, Bednarczyk D, Coleman J, Lee C, Palm J, Pak YA, Perloff ES, Reyner E, Balimane P, Brännström M, Chu X, Funk C, Guo A, Hanna I, Herédi-Szabó K, Hillgren K, Li L, Hollnack-Pusch E, Jamei M, *et al.* 2013. Variability in P-glycoprotein inhibitory potency (IC_{50}) using various *in vitro* experimental systems: implications for universal digoxin drug-drug interaction risk assessment decision criteria. Drug Metab Dispos. 41(7)1347-66.

Bleasby K, Castle JC, Roberts CJ, Cheng C, Bailey WJ, Sina JF, Kulkarni AV, Hafey MJ, Evers R, Johnson JM, Ulrich RG, Slatter JG. 2006. Expression profiles

of 50 xenobiotic transporter genes in humans and pre-clinical species: a resource for investigations into drug disposition. Xenobiotica. 36(10-11)963-88.

Bleicher KH, Böhm H, Müller K, Alanine AI. 2003. Hit and Lead Generation: Beyond High-Throughput Screening. Nat Rev Drug Discov. 2(5)369-78.

Böhm HJ, Schneider G. 2003. Protein ligand interactions. From molecular recognition to drug design. Wiley-VCH. Vol 19. pp.113-15.

Bonate P. Pharmacokinetic-Pharmacodynamic modeling and simulation. 2006. Springer Science. USA. pp. 56.

Borst P, Elferink RO. 2002. Mammalian ABC transporters in health and disease. Annu Rev Biochem. 71(1)537-92.

Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software. CRC Press. 1984.

Breiman L. 2001. Random forests. Machine learning. 45(1)5-32.

Broccatelli F, Carosati E, Neri A, Frosini M, Goracci L, Oprea TI, Cruciani G. 2011. A novel approach for predicting P-glycoprotein (ABCB1) inhibition using molecular interaction fields. J Med Chem. 54(6)1740-51.

Broggini M, Colombo T, Martini A, Donelli MG. 1980. Studies on the comparative distribution and biliary excretion of doxorubicin and 4'-epi-doxorubicin in mice and rats. Cancer treat rep. 64(8-9)897-904.

Bronchud MH, Foote M, Giaccone G, Olopade O. Principles of molecular oncology. Selecting the right target for cancer therapy. Humana press, Third edition. 2008.

Brown CD, Sayer R, Windass AS, Haslam IS, De Broe ME, D'Haese PC, Verhulst A. 2008. Characterisation of man tubular cell monolayers as a model of proximal tubular xenobiotic handling. 233(3)428-38.

Burckhardt G, Wolff NA. 2000. Structure of renal organic anion and cation transporters. Am J Physiol Renal Physiol. 278(6)853-66.

Burden FR. 1989. Molecular identification number for substructure searches. J Chem Inf Comput Sci. 29(1)225–227.

Burden FR. 1997. A chemically intuitive molecular index based on the eigenvalues of a modified adjacency matrix. Quant Struct-Act Rel. 16(4)309–314

Burton ME, Shaw LE, Schentag JJ, Evans WE. 2006. Applied pharmacokinetics and Pharmacodynamics. Principles of therapeutic drug monitoring. Lippincott Williams and Wilkins. pp. 132.

Carlson TJ, Fisher MB. 2008. Recent advances in high throughput screening for ADME properties. Comb Chem High Throughput Screen. 11(3) 258-64.

Cha SH, Sekine T, Fukushima JI, Kanai Y, Kobayashi Y, Goya T, Endou H. 2001. Identification and characterization of human organic anion transporter 3 expressing predominantly in the kidney. Mol Pharmacol. 59(5)1277-86.

Chan YL, Chou MH, Lin MF, Chen CF, Tsai TH. 2002. Determination and pharmacokinetic study of meropenem in rat bile using on-line microdialysis and liquid chromatography. J Chromatogr A. 961(1)119-24.

Chang C, Bahadduri PM, Polli JE, Swaan PW, Ekins S. 2006. Rapid Identification of P-glycoprotein Substrates and Inhibitors. Drug Met Dispos. 34(12)1976-84.

Chang C, Pang KS, Swaan PW, Ekins S. 2005. Comparative pharmacophore modeling of organic anion transporting polypeptides: a meta-analysis of rat Oatp1a1 and human OATP1B1. J Pharmacol Exp Ther. 314(2)533-41.

Chem med chem. 2007. Chemistry and drug discovery supporting information. Wiley-VCH verlag. Page 2.

ChemSpider (The free chemical database). Available from: http://www.chemspider.com (Accessed on 7 November 2012)

Chen L, Li Y, Yu H, Zhang L, Hou T. 2012. Computational Models for Predicting Substrates or Inhibitors of P-glycoprotein. Drug Discov Today. 17(7-8)343-51.

Chen Y, Cameron K, Guzman-Perez A, Perry D, Li D, Gao H. 2010. Structure-pharmacokinetic relationship of *in vivo* rat biliary excretion. Biopharm Drug Dispos. 31(1)82-90.

Chen ZS, Guo Y, Belinsky MG, Kotova E, Kruh GD. 2005. Transport of bile acids, sulphated steroids, estradiol 17-beta-D-glucuronide, and leukotriene C4 by human multidrug resistance proteins 8 (ABCC11). Moj Pharmacol. 67(2)545-57.

Cheng Y, Prusoff WH. 1973. Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem Pharmacol. 22(23)3099-108.

Chesnokova CA, Shaston CA, Agadkhanian NA. 2007. Atlas of Normal Physiology. Published by Medical News Agency.Chapter 5.

Chiba M, Jin L, Neway W, Vacca JP, Tata JR, Chapman K, Lin JH. 2001. P450 interaction with HIV protease inhibitors: relationship between metabolic stability, inhibitory potency, and P450 binding spectra. Drug Metab Disp. 29(1)1-3.

Chirico N, Gramatica P. 2012. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. J Chem Inf Model. 52(8)2044-58.

Choo EF, Leake B, Wandel C, Imamura H, Wood AJ, Wilkinson GR, Kim RB. 2000. Pharmacological inhibition of P-glycoprotein transport enhances the distribution of HIV-1 protease inhibitors into brain and testes. Drug Metab Dispos. 28(6)655-60.

Chu XY, Kato Y, Niinuma K, Sudo KI, Hakusui H, Sugiyama Y. 1997. Multispecific organic anion transporter is responsible for the biliary excretion of the camptothecin derivative irinotecan and its metabolites in rats. J Pharmacol Exp Ther. 281(1)304-14.

Chufan EE, Kapoor K, Sim HM, Singh S, Talele TT, Durell SR, Ambudkar SV. 2013. Multiple Transport-Active Binding Sites Are Available for a Single Substrate on Human P-Glycoprotein (ABCB1). PLoS One. 8(12)e82463.

Cihlar T, Lin DC, Pritchard JB, Fuller MD, Mendel DB, Sweet DH. 1999. The antiviral nucleotide analogs cidofovir and adefovir are novel substrates for human and rat renal organic anion transporter 1. Mol Pharmacol. 56(3)570-80.

Coleman MD. 2005. Human drug metabolism, an introduction. Appendix A. Method in drug metabolism. John Wiley and Sons Ltd. pp. 215-18.

Conseil G, Baubichon-Cortay H, Dayan G, Jault JM, Barron D, Di Pietro A. 1998. Flavonoids: a class of modulators with bifunctional interactions at vicinal ATP- and steroid-binding sites on mouse P-glycoprotein. Proc Natl Acad Sci USA. 95(17)9831-6.

Cook JA, Feng B, Fenner KS, Kempshall S, Liu R, Rotter C, Smith DA, Troutman MD, Ullah M, Lee CA. 2010. Refining the *in vitro* and *in vivo* critical parameters for P-glycoprotein, [I]/IC_{50} and [I2]/IC_{50}, that allow for the exclusion of drug candidates from clinical digoxin interaction studies. Mol Pharm. 7(2)398-411.

Copeland RA. 2005. Evaluation of enzyme inhibitors in drug discovery. Vol 46. Wiley Interscience. New Jersey.

Corbeil CR, Williams CI, Labute P. 2012. Variability in docking success rates due to dataset preparation. J Comput Aided Mol Des. 26(6)775-86.

Crespi CL, Stresser DM. 2000. Fluorometric screening for metabolism-based drug-drug interactions. J Pharmacol Toxicol Methods. 44(1)325-31.

Cronin MTD. Quantitive Structure-Activity relationships (QSARs). In: Puzyn T, Leszczynski J, Cronin MTD, editor. Recent advances in QSAR studies, methods and applications. New York: Vol 8. Springer; 2010. pp. 3-11.

Croop JM, Raymond M, Haber D, Devault A, Arceci RJ, Gros P, Housman DE. 1989. The three mouse multidrug resistance (mdr) genes are expressed in a tissue-specific manner in normal mouse tissues. Mol Cell Biol. 9(3)1346-50.

Crosignani A. 1996. Clinical pharmacokinetics of therapeutic bile acids. J Clin Pharmacokinet. 30(5)333-58.

Cruciani G, Crivori P, Carrupt PA, Testa B. 2000a. Molecular fields in quantitative structure–permeation relationships: the VolSurf approach. J Mol Struct THEOCHEM. 503(1–2)17–30.

Cruciani G, Pastor M, Guba W. 2000b. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. Eur J Pharm Sci. 11 (Suppl 2)29-39.

Cumming JG, Davis AM, Muresan S, Haeberlein M, Chen H. 2013. Chemical predictive modelling to improve compound quality. Nat Rev Drug Discov. 12(12)948-62.

Cummings MD, DesJarlais RL, Gibbs AC, Mohan V, Jaeger EP. 2005. Comparison of Automated Docking Programs as Virtual Screening Tools. J Med Chem. 48(4)962-76.

Cvetkovic M, Leake B, Fromm MF, Wilkinson GR, Kim R.B. 1999. OATP and P-glycoprotein transporters mediate the cellular uptake and excretion of fexofenadine. Drug Metab Dispos. 27(8)866-71.

Daniel H. 2004. Molecular and integrative physiology of intestinal peptide transport. Annu Rev Physiol. 66(1)361-384.

Dassa E, Bouige P. 2001. The ABC of ABCS: a phylogenetic and functional classification of ABC systems in living organisms. Res Microbiol. 152(3-4)211-29.

Davidson AL, Dassa E, Orelle C, Chen J. 2008. Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems. Microbiol Mol Biol Rev. 72(2)317-64.

Davis AM, Teague SJ. 1999. Hydrogen Bonding, Hydrophobic Interactions, and Failure of the Rigid Receptor Hypothesis. Angew Chem Int Ed. 38(6)736-749.

Dawson PA, Lan T, Rao A. 2009. Bile acid transporters. Journal of lipid research, Thematic review series: Bile acids. Vol 50. pp. 2340-57.

De Bruyn T, van Westen GJ, Ijzerman AP, Stieger B, de Witte P, Augustijns PF, Annaert PP. 2013. Structure-based identification of OATP1B1/3 inhibitors. Mol Pharmacol. 83(6)1257-67.

de Lannoy IA, Silverman M. 1992. The MDR1 gene product, P-glycoprotein, mediates the transport of the cardiac glycoside, digoxin. Biochem Biophys Res Commun. 189(1)551-7.

de Vree JM, Jacquemin E, Sturm E, Cresteil D, Bosma PJ, Aten J, Deleuze JF, Desrochers M, Burdelski M, Bernard O, Oude Elferink RP, Hadchouel M.1998. Mutations in the MDR3 gene cause progressive familial intrahepatic cholestasis. Proc Natl Acad Sci U S A. 95(1)282-7.

Dean MC. The Human ATP-Binding Cassette (ABC) Transporter Superfamily, First ed., Bethesda (MD), National Center for Biotechnology Information, New York, 2002.

Dearden JC, Cronin MTB. 2005. Quantitative structure-activity relationships (QSAR) in drug design. In: Smith H.J., (ed.) Smith and Williams' Introduction to the Principles of Drug Design and Action. 4 th Edition. Taylor and Francis, Boca Raton FL, USA, 185-209.

Dearden JC, Ghafourian T. 1999. Hydrogen bonding parameters for QSAR: comparison of indicator variables, hydrogen bond counts, molecular orbital and other parameters. J Chem Inf Comput Sci. 39(2)231–5.

De'ath G, Fabricius KE. 2000. Classification and Regression Trees: A powerful yet simple technique for ecological data analysis. Ecology. 81(11)3178-92.

Demel MA, Krämer O, Ettmayer P, Haaksma EE, Ecker GF. 2009. Predicting ligand interactions with ABC transporters in ADME. Chem Biodivers. 6(11)1960-9.

Denk GU, Soroka CJ, Takeyama Y, Chen YQ, Schuetz JD, Boyer JL. 2004. Multidrug resistance-associated protein 4 is up-regulated in liver but down-regulated in kidney in obstructive cholestasis in the rat. J Hepatol. 40(4)585-91.

Di Pietro A, Conseil G, Pe´rez-Victoria JM, Dayan G, Baubichon-Cortay H, Trompier D, Steinfels, Jault JM, de Wet H, Maitrejean M, Comte G, Boumendjel A, Mariotte AM, Dumontet C, McIntosh DB, Goffeau A, Castanys S, Gamarro F, Barron D. 2002. Modulation by flavonoids of cell multidrug resistance mediated by P-glycoprotein and related ABC transporters. Cell Mol Life Sci. 59(2)307–22.

Dimitrov S, Dimitrova G, Pavlov T, Dimitrova N, Patlewicz G, Niemela J, Mekenyan O. 2005. A stepwise approach for defining the applicability domain of SAR and QSAR models. J Chem Inf Model. 45(4)839-49.

DiPiro JT, Spruill WJ, Wade WE, Blouin RA, Pruemer JM. 2010. Concepts in clinical pharmacokinetics. Fifth edition. Published by American Society of Health-System Pharmacists.

Djulbegovic B, Hozo I, Ioannidis JP. 2014. Improving the drug development process: more not less randomized trials. JAMA. 311(4)355-6.

Doring F, Will J, Amashesh S, Clauss W, Ahlbrecth H, Daniel H. 1998. Minimal molecular determinants of substrates for recognition by the intestinal peptide transporter. J Biol Chem. 273(36)23211-18.

Doyle L, Ross DD. 2003. Multidrug resistance transporter from human MCF-7 breast cancer resistance protein BCRP (ABCG2). Oncogene. 22(47)7340-7358.

Doyle LA, Yang W, Abruzzo LV, Krogmann T, Gao Y, Rishi AK, Ross DD. 1998. A multidrug resistance transporter from human MCF-7 breast cancer cells. Proc Natl Acad Sci U S A. 95(26)15665-70.

Dror O, Shulman-Peleg A, Nussinov R, Wolfson HJ. 2004. Predicting molecular interactions *in silico*: I. A guide to pharmacophore identification and its applications to drug design. Curr Med Chem. 11(1)71-90.

Dudek AZ, Arodz T, Galvez J. 2006. Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review. Comb Chem High Throughput Screen. 9(3)213-228.

Eberl S, Renner B, Neubert A, Reisig M, Bachmakov I, König J, Dörje F, Mürdter TE, Ackermann A, Dormann H, Gassmann KG, Hahn EG, Zierhut S, Brune K, Fromm MF. 2007. Role of p-glycoprotein inhibition for drug interactions: evidence from *in vitro* and pharmacoepidemiological studies. Clin Pharmacokinet. 46(12)1039-49.

Eckford PD, Sharom FJ. 2009. ABC efflux pump-based resistance to chemotherapy drugs. Chem Rev.109(7)2989-3011.

Edwards JE, Alcorn J, Savolainen J, Anderson BD, McNamara PJ. 2005. Role of P-glycoprotein in distribution of nelfinavir across the blood-mammary tissue barrier and blood-brain barrier. Antimicrob Agents Chemother. 49(4)1626-28.

Ehrhardt C, Kim KJ. Drug absorption studies, in situ, *in vitro* and *in silico* models, in: T. Terada, K.I. Inui, Impact of drug transport proteins, Springer. New York, 2008, pp. 559-570.

Ekins S, Erickson J.A. 2002. A pharmacophore for human pregnane X receptor ligands, Drug Metab Dispos. 30 (1) 96-9.

Ekins S, Kim RB, Leake BF, Dantzig AH, Schuetz EG, Lan LB, Yasuda K, Shepard RL, Winter MA, Schuetz JD, Wikel JH, Wrighton SA. 2002a Three-dimensional quantitative structure-activity relationships of inhibitors of P-glycoprotein, Mol Pharmacol. 61(5)964-973.

Ekins S, Kim RB, Leake BF, Dantzig AH, Schuetz EG, Lan LB, Yasuda K, Shepard RL, Winter MA, Schuetz JD, Wikel JH, Wrighton SA. 2002b. Application of three-dimensional quantitative structure-activity relationships of P-glycoprotein inhibitors and substrates. Mol Pharmacol. 61(5):974-81.

Ekins S, Mestres J, Testa B. 2007. *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling. Br J Pharmacol. 152(1)9-20.

Elimrani I, Lahjouji K, Seidman E, Roy MJ, Mitchell GA, Qureshi I. 2003. Expression and localization of organic cation/carnitine transporter OCTN2 in Caco-2 cells. Am J Physiol Gastrointest Liver Physiol. 284(5)863-71.

El-Kattan A, Varma M. 2012. Oral Absorption, Intestinal Metabolism and Human Oral Bioavailability, Topics on Drug Metabolis. In: Paxton J (Ed.), Available from: http://www.intechopen.com/books/topics-on-drug-metabolism/oral-absorption-intestinalmetabolism-and-human-oral-bioavailability).

Enomoto A, Kimura H, Chairoungdua A, Shigeta Y, Jutabha P, Cha SH, Hosoyamada M, Takeda M, Sekine T, Igarashi T, Matsuo H, Kikuchi Y, Oda T, Ichida K, Hosoya T, Shimokata K, Niwa T, Kanai Y, Endou H. 2002. Molecular identification of a renal urate anion exchanger that regulates blood urate levels. Nature. 417(6887)447-52.

Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Perspect. 111(10)1361-75.

Eriksson UG, Dorani H, Karlsson J, Fritsch H, Hoffmann KJ, Olsson L, Sarich TC, Wall U, Schützer KM. 2006. Influence of erythromycin on the pharmacokinetics of ximelagatran may involve inhibition of P-glycoprotein-mediated excretion. Drug Metab Dispos. 34(5)775-82.

Ette EI, Williams PJ. 2007. Pharmacometrics. The science of quantitative pharmacology. Wiley & Sons publication. Print in USA. pp. 237.

Evanchik MJ, Allen D, Yoburn JC, Silverman JA, Hoch U. 2009. Metabolism of (+)-1,4-dihydro-7-(trans-3-methoxy-4-methylamino-1-pyrrolidinyl)-4-oxo-1-(2-thiazolyl)-1,8-naphthyridine-3-carboxylic acid (voreloxin; formerly SNS-595), a novel replication-dependent DNA-damaging agent. Drug Metab Dispos. 37(3)594-601.

Fahrig L, Brasch H, Iven H. 1989. Pharmacokinetics of methotrexate (MTX) and 7-hydroxymethotrexate (7-OH-MTX) in rats and evidence for the metabolism of MTX to 7-OH-MTX. Cancer Chemother Pharmacol. 23(3)156-60.

FDA Guidelines, available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm292362.pdf (Accessed on 4 March 2014).

Fei YJ, Kanai Y, Nussberger S, Ganapathy V, Leibach FH, Romero MF, Singh SK, Boron WF, Hediger MA. 1994. Expression cloning of a mammalian proton-coupled oligopeptide transporter. Nature. 368(6471)563-6.

Feitsma KG. 1989. Unequal disposition of enantiomers of the organic cation oxyphenonium in the rat isolated perfused liver. J Pharm Pharmacol. 41(1)27-31.

Fenner KS, Jones HM, Ullah M, Kempshall S, Dickins M, Lai Y, Morgan P, Barton HA. 2012. The evolution of the OATP hepatic uptake transport protein family in DMPK sciences: from obscure liver transporters to key determinants of hepatobiliary clearance. Xenobiotica. 42(1)28-45.

Ferlay J, Shin HR, Bray F, Forman D, Mathers CD, Parkin D. 2010. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int J Cancer. 2010 Dec 15;127(12):2893-917.

Fernández-Barrena MG, Monte MJ, Latasa MU, Uriarte I, Vicente E, Chang HC, Rodriguez-Ortigosa CM, Elferink RO, Berasain C, Marin JJ, Prieto J, Ávila MA. 2012. Lack of Abcc3 expression impairs bile-acid induced liver growth and delays hepatic regeneration after partial hepatectomy in mice. J Hepatol. 56(2)367-73.

Fortuna CG, Barresi V, Berellini G, Musumarra G. 2008. Design and synthesis of trans 2-(furan-2-yl)vinyl heteroaromatic iodides with antitumour activity. Bioorg Med Chem. 16(7)4150-9.

Foxwell BM, Mackie A, Ling V, Ryffel B. 1989. Identification of the multidrug resistance-related P-glycoprotein as a cyclosporine binding protein. Mol. Pharmacol. 36(4)543–546.

Friedman JH. 1991. Multivariate Adaptive Regression Splines. Ann. Stat. 19(1)1-67.

Fukuda H, Ohashi R, Tsuda-Tsukimoto M, Tamai I. 2008. Effect of plasma protein binding on *in vitro-in vivo* correlation of biliary excretion of drugs evaluated by sandwich-cultured rat hepatocytes. Drug Metab Dispos. 36(7)1275-82.

Funakoshi S, Murakami T, Yumoto R, Kiribayashi Y, Takano M. 2005. Role of organic anion transporting polypeptide 2 in pharmacokinetics of digoxin and beta-methyldigoxin in rats. J Pharm Sci. 94(6)1196-203.

Gad SC. 2005. Hand book of drug discovery. Introduction: Drug discovery in the 21st century. Wiley-Interscience. Hoboken, New Jersey. pp. 1-8.

Gandhi YA, Morris ME. 2012. Re-evaluation of a quantitative structure pharmacokinetic model for biliary excretion in rats. Drug Metab Dispos. 40(7)1259-62.

Gao B, Hagenbuch B, Kullak-Ublick GA, Benke D, Aguzzi A, Meier PJ. 2000. Organic anion-transporting polypeptides mediate transport of opioid peptides across blood-brain barrier. J Pharmacol Exp Ther. 294(1)73-9.

Ghafourian T, Barzegar-Jalali M, Dastmalchi S, Khavari-Khorasani T, Hakimiha N, Nokhodchi A. 2006. QSPR models for the prediction of apparent volume of distribution. Int J Pharm. 319(1-2)82-97.

Ghafourian T, Bozorgi AH. 2010. Estimation of drug solubility in water, PEG 400 and their binary mixtures using the molecular structures of solutes. Eur J Pharm Sci. 40(5)430-40.

Ghafourian T, Cronin MTD. 2006. The Effect of Variable Selection on the Non-linear Modelling of Oestrogen Receptor Binding. QSAR Comb Sci. 25(10)824-35.

Ghafourian T, Dearden JC. 2000. The use of atomic charges and orbital energies as hydrogen-bonding-donor parameters for QSAR studies: comparison of MNDO, AM1 and PM3 methods. J Pharm Pharmacol. 52(6)603-10.

Ghafourian T, Haji Agha Bozorgi A. 2010. Estimation of drug solubility in water, PEG 400 and their binary mixtures using the molecular structures of solutes. Eur J Pharma Sci. 40(5)430-40.

Giacomini KM, Huang SM, Tweedie DJ, Benet LZ, Brouwer KL, Chu X, Dahlin A, Evers R, Fischer V, Hillgren KM, Hoffmaster KA, Ishikawa T, Keppler D, Kim RB, Lee CA, Niemi M, Polli JW, Sugiyama Y, Swaan PW, Ware JA, Wright SH, Yee SW, Zamek-Gliszczynski MJ, Zhang L. 2010. Membrane transporters in drug development. International Transporter Consortium. Nat Rev Drug Discov. 9(3)215-36.

Gibson GG, Skett P. Introduction to drug metabolism. Chapter 4. Factors affecting drug metabolism: Internal factors. 2001. Third edition. Nelson Thornes Ltd. Cheltenham, UK. pp. 138-139.

Gil S, Saura R, Forestier F, Farinotti R. 2005. P-glycoprotein expression of the human placenta during pregnancy. Placenta. 26(2-3)268-70.

Glaeser H, Bailey DG, Dresser GK, Gregor JC, Schwarz UI, McGrath JS, Jolicoeur E, Lee W, Leake BF, Tirona RG, Kim RB. 2007. Intestinal drug transporter expression and the impact of grapefruit juice in humans. Clin Pharmacol Ther. 81(3)362-70.

Glaeser H, Kim RB. 2006. The relevance of transporters in determining drug disposition. In: Borchardt RT, Kerns EH, Hageman MJ, Thakker DR, Stevens JL (Eds.) Optimizing the “drug-like” properties of leads in drug discovery. pp. 423-60. New York. Springer.

Gleeson MP. 2008. Generation of a set of simple, interpretable ADMET rules of thumb. J Med Chem. 51(4)817-34.

Gohlke H, Klebe G. 2002. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Ed Engl. 41(15)2644-76.

Golbraikh A, Muratov E, Fourches D, Tropsha A. 2014. Dataset Modelability by QSAR. J Chem Inf Model. 54(1)1–4

Golbraikh A, Shen M, Xiao Z, Xiao Y, Lee K, Tropsha A. 2003. Rational selection of training and test sets for the development of validated QSAR models, Journal of Computer-Aided Molecular Design. 17(2-4)241–253.

Golbraikh A, Tropsha A. 2002. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J Comput Aided Mol Des. 16(5-6)357-69.

Gombar VK, Polli JW, Humphreys JE, Wring SA, Serabjit-Singh CS. 2004. Predicting P-glycoprotein substrates by a quantitative structure-activity relationship model. J Pharm Sci. 93(4)957-68.

Goodford PJ. 1985. A computational procedure for determining energetically favourable binding sites on biologically important macromolecules. J Med Chem. 28(7)849-57.

Gorboulev V, Ulzheimer JC, Akhoundova A, Ulzheimer-Teuber I, Karbach U, Quester S, Baumann C, Lang F, Busch AE, Koepsell H. 1997. Cloning and characterization of two human polyspecific organic cation transporters. DNA Cell Biol.16(7)871-81.

Gottesman MM, Ambudkar SV, Xia D. 2009. Structure of a multidrug transporter. Nat Biotechnol. 27(6)546-547.

Gottesman MM, Ambudkar SV. 2001. Overview: ABC transporter and human disease. J Bioenerg Biomembr. 33(6)453-456.

Gottesman MM, Fojo T, Bates SE. 2002. Multidrug resistance in cancer: role of ATP-dependent transporters. Nat Rev Cancer. 2(1)48-58.

Gottesman MM, Pastan I. 1988. The multidrug transporter, a double-edged sword. J Biol Chem. 263(25)12163-6.

Gottesman MM. 2002. Mechanisms of cancer drug resistance. Annu Rev Med. 53(1)615-27.

Gramatica P. 2011. A short history of QSAR evolution. Available from: http://www.qsarworld.com/ Temp_Fileupload/Shorthistoryofqsar.pdf (accessed Octoberber 10, 2012).

Gramatica P. 2013. On the Development and Validation of QSAR Models. In: Reisfeld B, Mayeno AN. Computational toxicology. Methods in molecular biology. Vol 2. Springer. New York. pp. 492-536.

Grundemann D, Gorboulev V, Gambaryan S, Veyhl M, Koepsell H. 1994. Drug excretion mediated by a new prototype of polyspecific transporter. Nature. 372(6506)549–52.

Guha R, Jurs PC. 2004. Development of linear, ensemble, and nonlinear models for the prediction and interpretation of the biological activity of a set of PDGFR inhibitors. J Chem Inf Comput Sci. 44(6)2179–89.

Guo Y, Kotova E, Chen ZS, Lee K, Hooper-Borge E, Belinsky MG, Kruh GD. 2003. MRP8 ATP-Binding cassette C11 (ABCC11), is a cyclic nucleotide efflux pump and a resistance factor for fluoropyrimidines 2' ,3' –dideoxycytidine and 9' –(2'-phosphonylmethoxyethyl) adenine. J Biol Chem. 278(32)29509-14.

Gutmann DA, Ward A, Urbatsch IL, Chang G, van Veen HW. 2010. Understanding polyspecificity of multidrug ABC transporters: closing in on the gaps in ABCB1. Trend Biochem Sci. 35(1)36-42.

Guvench O, MacKerell AD. 2008. Automated conformational energy fitting for force-field development. J Mol Model. 14(8)667-79.

Guyton AC, Hall JE. 2006. Textbook of medical physiology. Chapters 64 & 70. Eleventh edition. Elsevier Saunders. New York, USA. pp. 803-860.

Hagenbuch B, Gui C. 2008. Xenobiotic transporters of the human organic anion transporting polypeptides (OATP) family. Xenobiotica. 38(7-8)778-801.

Hagenbuch B, Meier PJ. 2003. The superfamily of organic anion transporting polypeptides. Biochim Biophys Acta. 1609(1)1–18.

Hagenbuch B, Meier PJ. 2004. Organic anion transporting polypeptides of the OATP/ SLC21 family: phylogenetic classification as OATP/ SLCO superfamily, new nomenclature and molecular/functional properties. Pflugers Arch. 447(5)653-65.

Hale TW, Llett KF. 2002. Drug Therapy and Breastfeeding. Chapter 2. Infant exposure to drugs. The Parthenon Publishing Group. New York, USA. pp. 15-20.

Halestrap AP, Meredith D. 2004. The SLC16 gene family-from monocarboxylate transporters (MCTs) to aromatic amino acid transporters and beyond. Eur J Physiol. 447(5)619-628.

Halestrap AP, Price NT. 1999. The proton-linked monocarboxylate transporter (MCT) family, structure, function and regulation. Bioche J. 343(2)281-99.

Hall LH, Kier LB. 1977. The nature of Structure-Activity Relationships and their relation to molecular connectivity. Eur J Med Chem 12(1)307-312.

Hall LH, Kier LB. 2007. The Molecular Connectivity Chi Indexes and kappa Shape Indexes in Structure-Property Modeling. Rev Comp Chem. 2(2)367-422.

Han H, De vrueh RL, Rhie JK, Covitz KM, Smith PM, Lee CP, Oh DM, Sadee W, Amidon G.L. 1998. 5'-Amino acid esters of antiviral nucleosides, acyclovir, and AZT are absorbed by the intestinal PEPT1 peptide transporter. Pharm Res. 15(8)1154–1159.

Han J, Kamber M. Data Mining: Concepts and Techniques, 2nd Edn., Morgan Kaufmann Publishers (Elsevier). 2006. USA.

Hansch C, Leo A, Mekapati SB, Kurup A. 2004. QSAR and ADME. Bioorg Med Chem. 12(12)3391-3400.

Haschek WM, Wallig AW, Rousseaux C. 2010. Fundamentals of toxicologic pathology. Second edition. London, England. Elsevier Inc. pp. 5-7.

Hawkins PC, Skillman AG, Nicholls A. 2007. Comparison of shape-matching and docking as virtual screening tools. J Med Chem. 50(1)74-82.

Hayouka Z, Hurevich M, Levin A, Benyamini H, Iosub A, Maes M, Shalev DE, Loyter A, Gilon C, Friedler A. 2010. Cyclic peptide inhibitors of HIV-1 integrase derived from the LEDGF/p75 protein. Bioorg Med Chem. 18(23)8388-95.

Hennessy M, Spiers JP. 2007. A Primer on the Mechanics of P-glycoprotein the Multidrug Transporter. Pharmacol Res. 55(1)1-15.

Herrera-Ruiz D, Wang Q, Gudmundsson OS, Cook TJ, Smith RL, Faria TN, Knipp GT. 2001. Spatial expression patterns of peptide transporters in the human and rat gastrointestinal tracts, Caco-2 *in vitro* cell culture, and multiple human tissues. AAPS PharmSci. 3(1)1-12.

Higgins CF, Gottesman MM. 1992. Is the multidrug transporter a flippase?. Trends Biochem Sci. 17(1)18-21.

Higgins CF. 2001. ABC transporters: physiology, structure and mechanism – an overview. Res Microbiol. 152(3-4)205-10.

Hilgendorf C, Ahlin G, Seithel A, Artursson P, Ungell AL, Karlsson J. 2007. Expression of thirty-six drug transporter genes in human intestine, liver, kidney, and organotypic cell lines. Drug Metab Dispos. 35(8)1333-40.

Hill T, Lewicki P. 2006. STATISTICS methods and applications. A comprehensive reference for science, industry and data mining. StatSoft Inc. 1st edition. Tulsa, USA. Chapter 12. Page 146.

Hillisch A, Hilgenfeld R. Modern Methods of Drug Discovery, In: Modern methods of drug discovery: An introduction. Giersiefen H, Hilgenfeld R, Hillisch A. 2003. Birkhäuser Verlag, Hamburg, Germany. pp. 2-18.

Hirano M, Maeda K, Shitara Y, Sugiyama Y. 2006. Drug-drug interaction between pitavastatin and various drugs via OATP1B1. Drug Metab Dispos. 34(7)1229-36.

Hirohashi T, Suzuki H, Ito K, Ogawa K, Kume K, Shimizu T, Sugiyama Y. 1998. Hepatic expression of multidrug resistance-associated protein-like proteins maintained in Eisai hyperbilirubinemic rats. Mol Pharmacol. 53(6)1068-75.

Hirom PC, Millburn P, Smith RL, Williams RT. 1972. Species variations in the threshold molecular-weight factor for the biliary excretion of organic anions. Biochem J. 129(5)1071–7.

Ho ES, Lin DC, Mendel DB, Cihlar T. 2000. Cytotoxicity of antiviral nucleotides adefovir and cidofovir is induced by the expression of human renal organic anion transporter 1. J Am Soc Nephrol. 11(3)383-93.

Ho RH, Tirona RG, Leake BF, Glaeser H, Lee W, Lemke CJ, Wang Y, Kim RB. 2006. Drug and bile acid transporters in rosuvastatin hepatic uptake: function, expression, and pharmacogenetics. Gastroenterology. 130(6)1793–1806.

Hong H, Xie Q, Ge W, Qian F, Fang H, Shi L, Su Z, Perkins R, Tong W. 2008. Mold2, Molecular Descriptors from 2D Structures for Chemoinformatics and toxicoinformatics. J Chem Info Model. 48(7)1337-1344.

Hooijberg JH, Broxterman HJ, Kool M, Assaraf YG, Peters GL, Noordhuis P, Scheper RJ, Borst P, Pinedo HM, Jansen G. 1999. Antifolate resistance mediated by the multidrug resistance proteins MRP1 and MRP2. Cancer Res. 59(11)2532-35.

Hosokawa S, Tagaya O, Mikami T, Nozaki Y, Kawaguchi a, Yamatsu K, Shamoto M. 1992. A new rat mutant with chronic conjugated hyperbilirubinemia and renal glomerular lesions. Lab Anim Sci. 42(1)27-34.

Hsiang B, Zhu Y, Wang Z, Wu Y, Sasseville V, Yang WP, Kirchgessner TG. 1999. A novel human hepatic organic anion transporting polypeptide (OATP2). Identification of a liver-specific human organic anion transporting polypeptide and identification of rat and human hydroxymethylglutaryl-CoA reductase inhibitor transporters. J Biol Chem. 274(52)37161-8.

Hughes RD, Millburn P, Williams RT. 1973. Biliary excretion of some diquaternary ammonium cations in the rats, guinea pig and rabbit. Biochem J. 136(4)979-84.

Hvidberg EF, Andreasen PB, Ranek L. 1974. Plasma half-life of phenylbutazone in patients with impaired hepatic function. J Clin Pharm Ther. 15(2)171-77.

Ismair MG, Stieger B, Cattori V, Hagenbuch B, Fried M, Meier PJ, Kullak-Ublick GA. 2001. Hepatic uptake of cholecystokinin octapeptide by organic anion-transporting polypeptides OATP4 and OATP8 of rat and human liver. Gastroenterology. 121(5)1185-90.

Israel M, Wilkinson PM, Pegg WJ, Frei E. 1978. Hepatobiliary metabolism and excretion of adriamycin and N-Trifluoroacetyladriamycin-14-valerate in the rat. Cancer Res. 38(2)365-70.

Itagaki S, Sugawara M, Kobayashi M, Miyazaki K, Iseki K. 2003. Mechanism of active secretion of phenolsulfonphthalein in the liver via Mrp2 (abcc2), an organic anion transporter. Drug Metab Pharmacokinet. 18(4)238-44.

Itoh T, Takemoto I, Itagaki S, Sasaki K, Hirano T, Iseki K. 2004. Biliary excretion of irinotecan and its metabolites. J Pharm Pharm. Sci. 7(1)13-8.

Iwata Y, Arisawa M, Hamada R, Kita Y, Mizutani MY, Tomioka N, Itai A, Miyamoto S. 2001. Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. J Med Chem. 44(11)1718-28.

Jacquemin E, Hagenbuch B, Stieger B, Wolkoff AW, Meier PJ. 1994. Expression cloning of a rat liver (Na^+)-independent organic anion transporter. Proc Natl Acad Sci USA. 91(1)133–37.

Jambhekar SS, Breen PJ. 2009. Basic pharmacokinetics. Pharmacetical press. First edition, London, England. Chapter one. pp.2-6.

Jansen PL, Peters WH, Lamers WH. 1985. Hereditary chronic conjugated hyperbilirubinemia in mutants rats caused by defective hepatic anion transport. Hepatology. 5(4)573-79.

Janku I. 1993. Physiological modelling of renal drug clearance. Eur J clin pharmacol. 44(6)513-19.

Jariyawat S, Sekine T, Takeda M, Apiwattanakul N, Kanai Y, Sophasan S, Endou H. 1999. The interaction and transport of beta-lactam antibiotics with the cloned rat renal organic anion transporter 1. J Pharmacol Exp Ther. 290(2)672-7.

Jedlitschky G, Leier I, Buchholz U, Barnouin K, Kurz G, Keppler D. 1996. Transport of glutathione, glucuronate, and sulphate conjugates by the MRP gene-encoded conjugate export pump. Cancer Res. 56(5)988-94.

Jemal A, Bray F, Center M, Ferlay J, Ward E, Forman D. 2011. Global cancer statistics. CA Cancer J Clin. 61(2)69-90.

Juliano RL, Ling V. 1976. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. Biochim Biophys Acta. 455(1)152-62.

Kakumoto M, Takara K, Sakaeda T, Tanigawara Y, Kita T, Okumura K. 2002. MDR1-mediated interaction of digoxin with antiarrhythmic or antianginal drugs. Biol Pharm Bull. 25(12)1604-7.

Kalliokoski A, Niemi M. 2009. Impact of OATP transporters on pharmacokinetics. Br J Pharmacol. 158(3)693-705.

Karelson M, Lobanov VS, Katritzky AR. 1996. Quantum-Chemical Descriptors in QSAR/QSPR Studies. Chem Rev. 96(3)1027-1044.

Karlgren M, Vildhede A, Norinder U, Wisniewski JR, Kimoto E, Lai Y, Haglund U, Artursson P. 2012a. Classification of inhibitors of hepatic organic anion transporting polypeptides (OATPs): influence of protein expression on drug-drug interactions. J Med Chem. 55(10)4740-63.

Karlgren M, Ahlin G, Bergström CA, Svensson R, Palm J, Artursson P. 2012b. *In vitro* and *in silico* strategies to identify OATP1B1 inhibitors and predict clinical drug-drug interactions. Pharm Res. 29(2) 411-26.

Kass GV. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. Applied Stat. 29(2)119-27.

Katoh M, Nakajima M, Yamazaki H, Yokoi T. 2001. Inhibitory effects of CYP3A4 substrates and their metabolites on P-glycoprotein-mediated transport. Eur J Pharm Sci. 12(4)505-13.

Katritzky AR, Petrukhin R, Tatham D, Basak S, Benfenati E, Karelson M, Maran U. 2001. Interpretation of quantitative structureproperty and -activity relationships. J Chem Inf Comput. Sci. 41(3)679-85.

Katzung BG, Trevor AJ, Masters S. 2004. Pharmacology examination and board review, eleventh edition, McGraw Hill publication. pp. 36.

Kemmerer JM, Rubio FA, McClain RM, Koechlin BA. 1979. Stereospecific assay and stereospecific disposition of racemic carprofen in rats. J Pharm Sci. 68(10)1274-80.

Keogh JP, Kunta JR. 2006. Development, validation and utility of an *in vitro* technique for assessment of potential clinical drug-drug interactions involving P-glycoprotein. Eur J Pharm Sci. 27(5)543-54.

Keogh JP. Membrane Transporters in Drug Development. In: Hawksworth GM, (Ed), Current Concepts in Drug Metabolism and Toxicology, of Advances in pharmacology, Vol 63. 2012. pp. 1-42.

Keppler D, König J, Nies AT. 2001. Conjugated export pumps of the multidrug resistance protein (MRP) family in liver. In: Arias IM, Boyer JL, Chisari FV, Fausto N, Schachter D, Shafritz DA, editors. The liver: Biology and pathobiology. New York: Lippincott Williams and Wilkins. pp. 373-382.

Keppler D, König J. 2000. Hepatic secretion of secretion of conjugated drugs and endogenous substances. Semin Liver Dis. 20(3)265-72.

Kerns EH, Di L. Drug-like properties: Concepts, structure, design and methods. 1st ed. London: Elsevier; 2008.

Kier LB, Hall LH. 1999. Molecular Structure Description: the Electrotopological State; Academic Press: San Diego.

Kim RB, Fromm MF, Wandel C, Leake B, Wood AJ, Roden DM. 1998. The Drug Transporter P-glycoprotein Limits Oral Absorption and Brain Entry of HIV-1 Protease Inhibitors. J Clin Invest. 101(2)289-294.

Kim RB. 2002. Drugs as P-glycoprotein substrates, inhibitors, and inducers. Drug Metab Rev. 34(1-2)47-54.

Kim RB. 2003. Organic anion-transporting polypeptide (OATP) transporter family and drug disposition. Eur J Clin Invest. 33(Suppl)21-5.

Kim WY, Benet LZ. 2004. P-glycoprotein (P-gp/MDR1)-mediated efflux of sex-steroid hormones and modulation of P-gp expression *in vitro*. Pharm Res. 21(7)1284-93.

Kirat D, Inoue H, Iwano H, Yokota H, Taniyama H, Kato S. 2007. Monocarboxylate transporter 1 (MCT1) in the liver of pre-ruminant and adult bovines. Vet J. 173(1)124-30.

Kitchen DB, Decornez H, Furr JR, Bajorath J. 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov. (11)935-49.

Klaassen CD. 1971. Biliary excretion in barbiturate. Br J Phamacol. 43(1)161-66.

Klein DM, Wright SH, Cherrington NJ. 2014. Localization of multidrug resistance-associated proteins along the blood-testis barrier in rat, macaque, and human testis. Drug Metab Dispos. 42(1)89-93.

Knutter I, Rubio-Aliaga I, Boll M, Hause G, Daniel H, Neubert K, Brandsch M. 2002. H^+-peptide cotransport in the human bile duct epithelium cell line SK-ChA-1. Am J Physiol Gastrointest Liver Physiol. 283(1)222-9.

Kobayashi D, Nozawa T, Imai K, Nezu J, Tsuji A, Tamai I. 2003. Involvement of human organic anion transporting polypeptide OATP-B (SLC21A9) in pH-dependent transport across intestinal apical membrane. J Pharmacol Exp Ther. 306(2)703-8.

Koenen A, Kroemer HK, Grube M, Meyer zu Schwabedissen HE. 2011. Current understanding of hepatic and intestinal OATP-mediated drug-drug interactions. Expert Rev Clin Pharmacol. 4(6)729-42.

Koepsell H, Gorboulev, Popp C, van Montfoort JE, Meier PJ, Arndt P, Volk C. Organic cation transporters in the sinusoidal membrance of hepatocytes. In: Matern S, Boyer JL, Keppler D, Meier-Abt PJ (editors). Hepatobiliary transport from bench to beside. London: Kluwer Academic; 2001. pp. 3-15.

Kohavi R, John GH. 1997. Wrappers for feature subset selection. Artif Intell. 97(1-2)273−324.

Kolhatkar V, Polli JE. 2010. Reliability of inhibition models to correctly identify type of inhibition. Pharm Res. 27(11)2433-45.

Kombo D, Tallapragada K, Jain R, Chewning J, Mazurov AA, Speake JD, Hauser TA, Toler S. 2013. 3D Molecular Descriptors important for clinical success. J. Chem Inf Model. 53(2)327–342.

Konig J, Cui Y, Nies AT, Keppler D. 2000a. A novel human organic anion transporting polypeptide localized to the basolateral hepatocyte membrane. Am J Physiol Gastrointest Liver Physiol. 278(1)156-64.

Konig J, Cui Y, Nies AT, Keppler D. 2000b. Localization and genomic organization of a new hepatocellular organic anion transporting polypeptide. J Biol Chem. 275(30)23161-8.

Kotani N, Maeda K, Watanabe T, Hiramatsu M, Gong LK, Bi YA, Takezawa T, Kusuhara H, Sugiyama Y. 2011. Culture period-dependent changes in the uptake of transporter substrates in sandwich-cultured rat and human hepatocytes. Drug Metab Dispos. 39(9)1503-10.

Kounnis V, Ioachim E, Svoboda M, Tzakos A, Sainis I, Thalhammer T, Steiner G, Briasoulis E. 2011. Expression of organic anion-transporting polypeptides 1B3, 1B1, and 1A2 in human pancreatic cancer reveals a new class of potential therapeutic targets. Onco Targets Ther. 4(1)27-32.

Krishna R, Mayer LD. 2000. Multidrug resistance (MDR) in cancer. Mechanisms, reversal using modulators of MDR and the role of MDR modulators in influencing the pharmacokinetics of anticancer drugs. Eur J Pharm Sci. 11(4)265-83.

Krishna R, McIntosh N, Riggs WK, Mayer LD. 1999. Doxorubicin Encapsulated in Sterically Stabilized Liposomes Exhibits Renal and Biliary Clearance Properties That Are Independent of Valspodar (PSC 833) under Conditions That Significantly Inhibit Nonencapsulated Drug Excretion. Clin Cancer Res. 5(10)2939-47.

Krishna R. 2004. Applications of Pharmacokinetic Principles in Drug Development. In: Sinz MW. Drug metabolism in preclinical development. Kluwer Academic publisher. New York. pp. 104-122.

Krogsgaard-Larsen P, Stromgaard K, Madsen U. Textbok of drug design and discovery. In: Jorgensen F.S., Kastrup J.S. Biostructure-based drug design. USA: CRC Press by Taylor & Francis Group; 2010. pp. 29-35 and 43.

Kruh GD, Guo Y, Hopper-Borge E, Belinsky MG, Chen ZS. 2007. ABCC10, ABCC11, and ABCC12. Pflugers Arch. 453(5)675-84.

Kubinyi H, Hamprecht FA, Mietzner T. 1998. Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices. J Med Chem. 41(14)2553-64.

Kullak-Ublick GA, Ismair MG, Stieger B, Landmann L, Huber R, Pizzagalli F, Fattinger K, Meier PJ, Hagenbuch B. 2001. Organic anion-transporting polypeptide B (OATP-B) and its functional comparison with three other OATPs of human liver. Gastroenterology. 120(2)525-33.

Kullak-Ublick GA, Stieger B, Hagenbuch B, Meier PJ. 2000. Hepatic Transport of Bile Salts. Semin Liver Dis. 20(2)273-92.

Kuncheva L, Whitaker C. 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, Machine. Learning. 51(1)181-207.

Kusuhara H, Sekine T, Utsunomiya-Tate N, Tsuda M, Kojima R, Cha SH, Sugiyama Y, Kanai Y, Endou H. 1999. Molecular cloning and characterization of a new multispecific organic anion transporter from rat brain. J Biol Chem. 274(19)13675-80.

Kusuhara H, Sugiyama Y. 2002. Role of transporters in the tissue-selective distribution and elimination of drugs: transporters in the liver, small intestine, brain and kidney. J Control Release. 78(1-3)43-54.

Kwon Y. Handbook of Essential Pharmacokinetics, Pharmacodynamics and Drug Metabolism for Industrial Scientists. Springer. 2001. Chapter 8. Metabolism. pp. 121-165.

Lahiri SR, Bolton N, Brown BL. 1970. Role of liver in Actinomycin-D metabolism. J Gasteroenter. 58(6)1021-8.

Lam JL, Okochi H, Huang Y, Benet LZ. 2006. *In vitro* and *in vivo* correlation of hepatic transporter effects on erythromycin metabolism: characterizing the importance of transporter-enzyme interplay. Drug Metab Dispos. 34(8)1336-44.

Lan LB, Ayesh S, Lyubimov E, Pashinsky I, Stein WD. 1996. Kinetic parameters for reversal of the multidrug pump as measured for drug accumulation and cell killing. Cancer Chemother Pharmacol. 38(2)181-90.

Lau YY, Okochi H, Huang Y, Benet LZ. 2006. Multiple transporters affect the disposition of atorvastatin and its two active hydroxy metabolites: application of *in vitro* and ex situ systems. J Pharmacol Exp Ther. 316(2)762-71.

Leabman MK, Huang CC, DeYoung J, Carlson EJ, Taylor TR, De la cruz M, Johns SJ, Stryke D, Kawamoto M, Urban TJ. 2003. Natural variation in human membrance transporter genes reveals evolutionary and functional constraints. Proc Natl Acad Sci USA. 100(10)5896-5901.

Leach A, Gillet V. 2003. An introduction to chemoinformatics. Kluwer Academic Publisher. Dordrecht, Netherlands. pp. 53-75.

Lee CA. Assessment of P-glycoprotein inhibitory potency (IC_{50}) variability in various *in vitro* experimental systems. AAPS Workshop on Drug Transporters in ADME: from the Bench to the Bedside; Bethesda, MD. 2011.

Lee JH, Lee MG. 2007. Effects of acute renal failure on the pharmacokinetics of telithromycin in rats: Negligible effects of increase in CYP3A1 on the metabolism of telithromycin. J Biopharm Drug Dispos. 28(4)157-166.

Leibach FH, Ganapathy V. 1996. Peptide transport in the intestine and the kidney. Annu Rev Nutr. 16(1)99-119.

Leonard JT, Roy K. 2008. Exploring molecular shape analysis of styrylquinoline derivatives as HIV-1 integrase inhibitors. Eur J Med Chem. 43(1)81-92.

Leslie EM, Deeley RG, Cole SP. 2009. Multidrug resistance proteins: role of P-glycoprotein, MRP1, MRP2, and BCRP (ABCG2) in tissue defense. Toxicol App Pharm. 204(3)216-37.

Leuthold S, Hagenbuch B, Mohebbi N, Wagner CA, Meier PJ, Stieger B. 2009. Mechanisms of pH-gradient driven transport mediated by organic anion polypeptide transporters. Am J Physiol Cell Physiol. 296(3)570-82.

Leveque D, Jehl F. 1995. P-glycoprotein and pharmacokinetics. Anticancer Res. 15(2)231-6.

Lewicki P, Hill S. Statistics, methods and applications, a comprehensive reference for science, industry and data mining. 1st ed. USA: StatSoft Inc; 2006.

Lewis DF, Ito Y. 2010. Human CYPs involved in drug metabolism: structures, substrates and binding affinities. Expert opinion on drug metabolism and toxicology. 6(6)661-674.

Li AP. Drug-drug interactions in pharmaceutical development. *In vitro* evaluation of metabolic drug-drug interactions: Concepts and practice. Canada: John Wiley and Sons; 2008. pp. 17-22.

Li F, Hong L, Mau CI, Chan R, Hendricks T, Dvorak C, Yee C, Harris J, Alfredson T. 2006. Transport of levovirin prodrugs in the human intestinal Caco-2 cell line. J Pharm Sci. 95(6)1318–25.

Li M, Yuan H, Li N, Song G, Zheng Y, Baratta M, Hua F, Thurston A, Wang J, Lai Y. 2008. Identification of interspecies difference in efflux transporters of hepatocytes from dog, rat, monkey and human. Eur J Pharm Sci. 35(1-2)114-26.

Liang R, Fei YJ, Prasad PD, Ramamoorthy S, Han H, Yang-Feng TL, Hediger MA, Ganapathy V, Leibach FH. 1995. Human intestinal H^+/peptide cotransporter: cloning, functional expression, and chromosomal localization. J Biol Chem. 270(12)6456-63.

Lickteig AJ, Xingguo C, Augustine LM, Klaassen CD, Cherrington NJ. 2008. Tissue distribution, ontogeny and induction of the transporters Multidrug and toxin extrusion (MATE) 1 and MATE2 mRNA expression levels in mice. Life Sci. 83(1-2)59-64.

Lin JH, Yamazaki M. 2003. Role of P-glycoprotein in pharmacokinetics: clinical implications. Clin Pharmacokinet. 42(1)59-98.

Lin JH. 2003. Drug-drug interaction mediated by inhibition and induction of P-glycoprotein. Adv Drug Deliv Rev. 55(1)53-81.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 46(1-3)3-26.

Lipinski CA, Lombardo F, Dominy BW, FeeneyPJ. 1997. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. Adv Drug Deliv Rev. 23(1-3)3-25.

Lipkowitz KB, Boyd DB. 2002. Reviews in computational chemistry, Vol 18. New Jersey: Wiley-VCH.

Lippmann ES, Azarin SM, Kay JE, Nessler RA, Wilson HK, Al-Ahmad A, Palecek SP, Shusta EV. 2012. Derivation of blood-brain barrier endothelial cells from human pluripotent stem cells. Nat Biotechnol. 30(8)783-91.

Lips KS, Volk C, Schmitt BM, Pfeil U, Arndt P, Miska D, Ermert L, Kummer W, Koepsell H. 2005. Polyspecific cation transporters mediate luminal release of acetylcholine from bronchial epithelium. Am J Respir Cell Mol Biol. 33(1) 79-88.

Liu W, Liang R, Ramamoorthy S, Fei YJ, Ganapathy ME, Hediger MA, Ganapathy V, Leibach FH. 1995. Molecular cloning of PEPT 2, a new member of the H^+/peptide cotransporter family, from human kidney. Biochim Biophys Acta. 1235(2)461-6.

Livingston DJ. Building QSAR models: A practical guide, Predicting Chemical Toxicity and Fate, In: MTD Cronin., CRC Press, USA, pp. 151-170. 2004.

Loo TW, Bartlett MC, Clarke DM. 2003. Substrate-induced conformational changes in the transmembrane segments of human P-glycoprotein. Direct evidence for the substrate-induced fit mechanism for drug binding. J Biol Chem. 278(16)13603-6.

Loo TW, Bartlett MC, Clarke DM. 2009. Identification of Residues in the Drug Translocation Pathway of the Human Multidrug Resistance P-glycoprotein by Arginine Mutagenesis. J Biol Chem. 284(36)24074-87.

Loscher W, Potschka H. 2005. Blood-brain barrier active efflux transporters: ATP-binding cassette gene family. NeuroRx. 2(1)86-98.

Löschmann N. Michaelis M. Rothweiler F. Zehner R. Cinatl J. Voges Y. Sharifi M, Riecken K, Meyer J, von Deimling A, Fichtner I, Ghafourian T, Westermann F, Cinatl J Jr. 2013. Testing of SNS-032 in a panel of human neuroblastoma cell lines with acquired resistance to a broad range of drugs. Transl Oncol. 6(6) 685-696.

Lu L, Leonessa F, Clarke R, Wainer IW. 2001. Competitive and allosteric interactions in ligand binding to P-glycoprotein as observed on an immobilized P-glycoprotein liquid chromatographic stationary phase. Mol Pharmacol. 59(1)62-8.

Lu R, Kanai N, Bao Y, Schuster VL. 1996. Cloning, *in vitro* expression, and tissue distribution of a human prostaglandin transporter cDNA(hPGT). J Clin Invest. 98(5)1142-9.

Lumen AA, Acharya P, Polli JW, Ayrton A, Ellens H, Bentz J. 2010. If the KI is defined by the free energy of binding to P-glycoprotein, which kinetic parameters define the IC50 for the Madin-Darby canine kidney II cell line overexpressing human multidrug resistance 1 confluent cell monolayer? Drug Metab Dispos.38(2)260-9.

Luo FR, Paranjpe PV, Guo A, Rubin E, Sinko P. 2002. Intestinal transport of irinotecan in Caco-2 cells and MDCK II cells overexpressing efflux transporters P-gp, cMOAT, and MRP1. Drug Metab Dispos. 30(7)763-70.

Luo G, Johnson S, Hsueh M, Zheng J, Hong C, Xin B, Chong S, He K, Harper TW. 2010. *In silico* prediction of biliary excretion of drugs in rats based on physicochemical properties. Drug Metab Dispos. 38(3)422-30.

Luscombe DK, Nicholis PJ. 1998. Processes of drug handling by the body. In: Smith HJ, Williams H, editors. Introduction to the principles of drug design and action. 3rd ed. Amsterdam: Harwood academic publishers. pp. 24-25.

Madan A, Usuki E, Burton L, Ogilvie B, Parkinson A. 2002. *In vitro* approaches for studying the inhibition of drug metabolism enzymes and the identifying the drug metabolism enzymes responsible for the metabolism of drugs. In: Rodrigues AA. Drug-Drug interactions. 1st ed. Marcel Dekker, New York. pp. 217-94.

Madon J, Hagenbuch B, Landmann L, Meier PJ, Stieger B. 2000. Transport function and hepatocellular localization of mrp6 in rat liver. Mol Pharmacol. 57(3)634-41.

Maeda K, Shitara Y, Horie T, Sugiyama Y. Web-based database as a tool to examine drug-drug interactions involving transporters. In: Pang SK, Rodrigues DA, Raimund MP, editors. Enzyme and transporter-based drug-drug interactions. Progress and future challenges. London: Springer; 2010. pp. 387-414.

Maki N, Moitra K, Ghosh P, Dey S. 2006. Allosteric modulation bypasses the requirement for ATP hydrolysis in regenerating low affinity transition state conformation of human P-glycoprotein. J Biol Chem. 281(16)10769-77.

Malik A, Singh H, Andrabi M, Husain SA, Ahmad S. 2006. Databases and QSAR for Cancer Research. Cancer Info. 5(2)99-111.

Malmo J, Sandvig A, Vårum KM, Strand SP. 2013. Nanoparticle mediated P-glycoprotein silencing for improved drug delivery across the blood-brain barrier: a siRNA-chitosan approach. PLoS One. 8(1)54182-9.

Mao Q, Unadkat JD. 2005. Role of the breast cancer resistance protein (ABCG2) in drug transport. AAPS J.7(1)118-33.

Marion TL, Perry CH, St Claire RL 3rd, Yue W, Brouwer KL. 2011. Differential disposition of chenodeoxycholic acid versus taurocholic acid in response to acute troglitazone exposure in rat hepatocytes. Toxicol Sci. 120(2)371-80.

Martin C, Berridge G, Higgins CF, Mistry P, Charlton P, Callaghan R. 2000. Communication between multiple drug binding sites on P-glycoprotein. Mol Pharmacol. 58(3)624-32.

Massey PR, Fojo T, Bates SE. Handbook of Anticancer Pharmacokinetics and Pharmacodynamics. In: Rudek MA, Chau CH, Figg WD, McLeod HL (editors). ABC Transporters: Involvement in Multidrug Resistance and Drug Disposition. London: Springer; pp. 373-400. 2014.

Masuda S, Terada T, Yonezawa A, Tanihara Y, Kishimoto K, Katsura K, Ogawa O, Inui K. 2006. Identification and functional characterization of a new human kidney-specific H^+/organic cation antiporter, kidney-specific multidrug and toxin extrusion 2. J Am Soc Nephrol. 17(8)2127–35.

Matern S, Boyer JL, Keppler D, Meier-Abt PJ. 2001. Hepatobiliary transport from bench to beside. Chapter: Organic cation transporters. Published by Kluwer Academic Publisher. pp. 10-13.

Matsson P, Englund G, Ahlin G, Bergström CA, Norinder U, Artursson P. 2007. A global drug inhibition pattern for the human ATP-binding cassette transporter breast cancer resistance protein (ABCG2). J Pharmacol Exp Ther. 323(1)19-30.

Matsson P, Pedersen JM, Norinder U, Bergström CAS, Artursson P. 2009. Identification of novel specific and general inhibitors of the three major human ATP-binding cassette transporters P-gp, BCRP, and MRP2 among registered drugs. Pharm Res. 26(8)1816-31.

Matsushima S, Maeda K, Kondo C, Hirano M, Sasaki M, Suzuki H, Sugiyama Y. 2005. Identification of the hepatic efflux transporters of organic anions using double-transfected Madin-Darby canine kidney II cells expressing human organic anion-transporting polypeptide 1B1 (OATP1B1)/multidrug resistance-associated

protein 2, OATP1B1/multidrug resistance 1, and OATP1B1/breast cancer resistance protein. J Pharmacol Exp Ther. 314(3)1059-67.

Matsushita H, Suzuki H, Sugiyama Y, Sawada Y, Iga T, Kawaguchi Y, Hanano M. 1992. Effect of benzylpenicillin on the disposition of cefodizime in rats: no net effect on total clearance due to decreased hepatobiliary clearance and increased renal clearance. J Pharmacol Exp Ther. 260(2)499-504.

Matsumoto K, Nakamura T. 1992. Hepatocyte growth factor: molecular structure, roles in liver regeneration, and other biological functions. Crit Rev Oncogen. 3(1-2)27-54

McAleer MA, Breen MA, White NL, Matthews N. 1999. ABC11 (also known as MOAT-C and MRP5), a member of the ABC family of proteins, has anion transporter activity but dose not confer multidrug resistance when overexpressed in human embryonic kidney 239 cells. J Biol Chem. 274(33)23541-58.

McDevitt CA, Callaaghan R. 2007. How can we best use structural information on P-glycoprotein to design inhibitors?. Pharmacol Ther. 113(2)429-41.

Meier-Abt F, Mokrab Y, Mizuguchi K. 2005. Organic anion transporting polypeptides of the OATP/SLCO superfamily: identification of new members in nonmammalian species, comparative modeling and a potential transport mode. J Membr Biol. 208(3)213-27.

Melaine N, Lienard MO, Dorval I, Le Goascogne C, Lejeune H, Jegou B. 2002. Multidrug resistance genes and P-glycoprotein in the testis of the rat, mouse, guinea pig and human. Biol Reprod. 67(6)1699-707.

Merino G, van Herwaarden AE, Wagenaar E, Jonker JW, Schinkel AH. 2005a. Sex-dependent expression and activity of the ATP-binding cassette transporter breast cancer resistance protein (BCRP/ABCG2) in liver. Mol Pharmacol. 67(5)1765-71.

Merino G, Jonker JW, Wagenaar E, van Herwaarden AE, Schinkel AH. 2005b. The breast cancer resistance protein (BCRP/ABCG2) affects pharmacokinetics,

hepatobiliary excretion, and milk secretion of the antibiotic nitrofurantoin. Mol Pharmacol. 67(5)1758-64.

Michaelis M, Rothweiler F, Nerreter T, Sharifi M, Ghafourian T, Cinatl J. 2014. Karanjin interferes with ABCB1, ABCC1, and ABCG2. J Pharm Pharmaceut Sci (US). 17(1)92-105.

Mihalic Z, Nikolic S, Trinajstic N. 1992. Comparative study of molecular descriptors derived from the distance matrix. J Chem Inf Comput Sci. 32(1)28-37

Mikkaichi T, Suzuki T, Tanemoto M, Ito S, Abe T. 2004. The organic anion transporter (OATP) family. Drug Metab Pharmacokinet. 19(3)171-9.

Millburn R, Smith RL, Williams RT. 1967. Biliary excretion of foreign compounds. Biochem J. 105(3)1275-81.

MOE Help File, 2012. Available at: https://www.chemcomp.com/MOE-Cheminformatics_and_QSAR.htm (Last accessed on 10 January 2014)

Mohri K, Okada K, Benet LZ. 2005. Stereoselective taurine conjugation of (R)-benoxaprofen enantiomer in rats: *in vivo* and *in vitro* studies using rat hepatic mitochondria and microsomes. Pharm Res. 22(1)79-85.

Morgan RE, Trauner M, van Staden CJ, Lee PH, Ramachandran B, Eschenberg M, Afshari CA, Hamadeh HK. 2010. Interference with bile salt export pump function is a susceptibility factor for human liver injury in drug development. Toxicol Sci. 118(2)485-500.

Morikawa A, Goto Y, Suzuki H, Hirohashi T, Sugiyama Y. 2000. Biliary excretion of 17beta-estradiol 17beta-D-glucuronide is predominantly mediated by cMOAT/MRP2. Pharm Res. 17(5)546-52.

Morita Y, Kataoka A, Shiota S, Mizushima T, Tsuchiya T. 2000. NorM of vibrio parahaemolyticus is an Na(+)-driven multidrug efflux pump. J Bacteriol. 182(23)6694-7.

Motohashi H, Inui K. 2013. Organic cation transporter OCTs (SLC22) and MATEs (SLC47) in the human kidney. AAPS J. 15(2)581-8.

Motohashi H, Nakao Y, Masuda S, Katsura T, Kamba T, Ogawa O, Inui K. 2013. Precise comparison of protein localization among OCT, OAT, and MATE in human kidney. J Pharm Sci. 102(9)3302-8.

Motohashi H, Sakurai Y, Saito H, Masuda S, Urakami Y, Goto M, Fukatsu A, Ogawa O, Inui K. 2002. Gene expression levels and immunolocalization of organic ion transporters in the human kidney. J Am Soc Nephrol. 13(4)866–74.

Muller J, Lips KS, Metzner L, Neubert RH, Koepsell H, Brandsch M. 2005. Drug specificity and intestinal membrane localization of human organic cation transporters (OCT). Biochem Pharmacol. 70(12)1851-60.

Murakawa T, Sakamoto H, Fukada S, Konishi T, Nishida M. 1982. Pharmacokinetics of fosmidomycin, a new phosphonic acid antibiotic. Antimicrob Agents Chemother. 21(2)224-30.

Nakanishi T, Shibue Y, Fukuyama Y, Yoshida K, Fukuda H, Shirasaka Y, Tamai I. 2011. Quantitative time-lapse imaging-based analysis of drug-drug interaction mediated by hepatobiliary transporter, multidrug resistance-associated protein 2, in sandwich-cultured rat hepatocytes. Drug Metab Dispos. 39(6)984-91.

Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V. 2010. Advances in computational methods to predict the biological activity of compounds. Expert Opin Drug Discov. 5(7)633-54.

Nassar AF, Hollenberg PF, Scatina J. 2009. Drug Metabolism handbook. Concepts and applications. In: Davis C.D., Rodrigues A.D. An introduction to metabolic reaction phenotyping. John Wiley and Sons Inc. New Jersey. First edition. pp. 391-446.

Neef C, Keulemans KT, Meijer DK. 1984. Hepatic uptake and biliary excretion of organic cations-I. Characterization of three new model compounds. Biochem Pharmacol. 33(24)3977-90.

Netzeva TI, Worth A, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Kahn S, Klopman G, Marchant C.A. *et al.* 2005. Current status of methods for defining the applicability domain of (quantitative) structure-activity

relationships. The report and recommendations of ECVAM Workshop 52. Altern Lab Anim. 33(2)155–73.

Neuhoff S, Langguth P, Dressler C, Andersson TB, Regårdh CG, Spahn-Langguth H. 2000. Affinities at the verapamil binding site of MDR1-encoded P-glycoprotein: drugs and analogs, stereoisomers and metabolites. Int J Clin Pharmacol Ther. 38(4)168-79.

Neuvonen PJ, Niemi M, Backman JT. 2006. Drug interactions with lipid-lowering drugs: mechanisms and clinical relevance. Clin Pharmacol Ther. 80(6)565-81.

Newby D, Freitas AA, Ghafourian T. 2013a. Pre-processing feature selection for improved C&RT models for oral absorption. J Chem Inf Model. 53(10)2730-42.

Newby D, Freitas AA, Ghafourian T. 2013b. Coping with unbalanced class datasets in oral absorption models. J Chem Inf Model. 53(2)461-74.

Ni Z, Bikadi Z, Rosenberg MF, Mao Q. 2010. Structure and function of the human breast cancer resistance protein (BCRP/ABCG2). Curr Drug Metab. 11(7)603-17.

Nicolle E, Boumendjel A, Macalou S, Genoux E, Ahmed-Belkacem A, Carrupt PA, Di Pietro A. 2009. QSAR analysis and molecular modeling of ABCG2-specific inhibitors. Adv Drug Deliv Rev. 61(1)34-46.

Niemi M, Pasanen MK, Neuvonen PJ. 2011. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. Pharmacol Rev. 63(1)157-81.

Nies AT, Koepsell H, Damme K, Schwab M. 2011. Organic cation transporters (OCTs, MATEs), *in vitro* and *in vivo* evidence for the importance in drug therapy. Handb Exp Pharmacol. 201(1)105-67.

Nies AT, Koepsell H, Winter S, Burk O, Klein K, Kerb R, Zanger UM, Keppler D, Schwab M, Schaeffeler E. 2009. Expression of organic cation transporters OCT1 (SLC22A1) and OCT3 (SLC22A3) is affected by genetic factors and cholestasis in human liver. Hepatol. 50(4)1227-40.

Nies AT, Rius M, Keppler D. 2007. Multidrug Resistance Proteins of the ABCC Subfamily. In: You, G. and Morris, M. eds. 2007. Drug Transporters: Molecular Characterization and Role in Drug Disposition. New Jersey: John Wiley & Sons, Inc, pp. 263-318.

Niinuma K, Kato Y, Suzuki H, Tyson CA, Weizer V, Dabbs JE, Froehlich R, Green CE, Sugiyama Y. 1999. Primary active transport of organic anions on bile canalicular membrane in humans. Am J Physiol. 276 (5 Pt 1)1153-64.

Noguchi K, Kawahara H, Kaji A, Katayama K, Mitsuhashi J, Sugimoto Y. 2009. Substrate-dependent bidirectional modulation of P-glycoprotein-mediated drug resistance by erlotinib. Cancer Sci. 100(9)1701-7.

Norinder U. 2003. Support vector machine models in drug design: applications to drug transport processes and QSAR using simplex optimisations and variable selection. Neurocomputing. 55(1-2)337-46.

Nozawa T, Minami H, Sugiura S, Tsuji A, Tamai I. 2005. Role of organic anion transporter OATP1B1 (OATP-C) in hepatic uptake of irinotecan and its active metabolite, 7-ethyl-10-hydroxycamptothecin: *in vitro* evidence and effect of single nucleotide polymorphisms. Drug Metab Dispos. 33(3)434-9.

Nozawa T, Tamai I, Sai Y, Nezu J, Tsuji A. 2003. Contribution of organic anion transporting polypeptide OATP-C to hepatic elimination of the opioid pentapeptide analogue (D-Ala2, D-Leu5)-enkephalin. J Pharm Pharmacol. 55(7)1013-20.

Obach RS, Lombardo F, Waters NJ. 2008. Trend analysis of a database of intravenous pharmacokinetic parameters in humans for 670 drug compounds. Drug Metab Dispos. 36(7)1385-405.

Obach RS, Nedderman AN, Smith DA. 2012. Radiolabelled mass-balance excretion and metabolism studies in laboratory animals: are they still necessary?. Xenobiotica. 42(1)46–56.

Ohta KY, Inoue K, Hayashi Y, Yuasa H. 2006. Molecular identification and functional characterization of rat multidrug and toxin extrusion type transporter 1 as an organic cation/H^+ antiporter in the kidney. Drug Metab Dispos. 34(11)1868-74.

Okuda M, Saito H, Urakami Y, Takano M, Inui K. 1996. cDNA cloning and functional expression of a novel rat kidney organic cation transporter, OCT2. Biochem Biophys Res Commun. 224(2)500–7.

Oostendorp RL, Buckle T, Beijnen JH, van Tellingen O, Schellens JH. 2008. The effect of P-gp (Mdr1a/1b), BCRP (Bcrp1) and P-gp/BCRP inhibitors on the *in vivo* absorption, distribution, metabolism and excretion of imatinib. Invest New Drugs. 27(1)31-40.

Oprea TI. 2000. Property Distribution of Drug-Related Chemical Databases; J. Comp. Aid. Mol. Des. 14(3)251–264.

O'Reilly WJ, Pitt PA, Ryan AJ. 1971. Pharmacokinetic model for the successive demethylation and biliary secretion of methyl orange in the rat. Br J Pharmacol. 43(1)167-179.

Oza AM. 2002. Clinical development of P glycoprotein modulators in oncology. Novartis Found Symp. 243(1)103-18.

Paintaud G, Bechtel Y, Brientini MP, Miguet JP. 1996. Effects of liver diseases on drug metabolism. Therapie. 51(4)384-9.

Pajeva IK, Globisch C, Wiese M. 2009. Combined pharmacophore modeling, docking, and 3D QSAR studies of ABCB1 and ABCC1 transporter inhibitors. ChemMedChem. 4(11)1883-96.

Pandit NK. Introduction to the pharmaceutical sciences. First edition, 2007. Lippincott Williams and Wilkins. New York.

Patel H, Berge WT, Cornin MTD. 2002. Quantitative structure-activity relationships (QSAR) for the prediction of skin permeation of exogenous chemicals. Chemoshere. 48(6)603-13.

Patric JS. 2006. Martin’s Physical pharmacy and pharmaceutical science. Lippincot Williams and Wilkins. Fifth edition. pp. 392-4.

Pauli-Magnus C, von Richter O, Burk O, Ziegler A, Mettang T, Eichelbaum M, Fromm MF. 2000. Characterization of the major metabolites of verapamil as substrates and inhibitors of P-glycoprotein. J Pharmacol Exp Ther. 293(2)376-82.

Pearlman R S. Smith K M. Novel Software Tools for Chemical Diversity, In: Kubinyi H, Martin Y, Folkers G, Eds; 3D-QSAR and Drug Design: Recent Advances; Kluwer Academic: Dordrecht, Netherlands. 1997. pp 339-53.

Perloff MD, Von Moltke LL, Marchand JE, Greenblatt DJ. 2001. Ritonavir induces P-glycoprotein expression, multidrug resistance-associated protein (MRP1) expression, and drug transporter-mediated activity in a human intestinal cell line. J Pharm Sci. 90(11)1829-37.

Petri N, Tannergren C, Rungstad D, Lennernäs H. 2004. Transport characteristics of fexofenadine in the Caco-2 cell model. Pharm Res.21(8)1398-404.

Pfeifer ND, Hardwick RN, Brouwer KL. 2014. Role of hepatic efflux transporters in regulating systemic and hepatocyte exposure to xenobiotics. Annu Rev Pharmacol Toxicol. 54(1)509-35.

Piscitelli SC, Rodvold KA. 2005. Drug interactions in infectious diseases. In: Fish D.N., editor. Non-HIV antiviral agents Humana press. New York, USA. Second edition. pp. 498.

Plusquellec Y, Arnaud R, Saivin S. 1998. Enterohepatic recirculation of the new antihypertensive drug UP 269-9 in humans: a possible model to account for multiple plasma peaks. Arzneimittel-Forschung. 48(2)138-144.

Pocock G, Richards CD. 2009. The human body. An introduction for the biomedical and health sciences. The liver and gall bladder. Oxford university press. First edition. London, UK. Chapter 25. pp. 532-42.

Poongavanam V, Haider N, Ecker GF. 2012. Fingerprint-based *in silico* models for the prediction of P-glycoprotein substrates and inhibitors. Bioorg Med Chem. 20(18)5388–95.

Proost JH, Roggeveld J, Wierda JM, Meijer DK. 1997. Relationship between chemical structure and physicochemical properties of series of bulky organic cations and their hepatic uptake and biliary excretion rates. J Pharmacol Exp Ther. 282(2)715-26.

Prueksaritanont T, Xu X, Deluna P, Yamazaki M, Lin JH. 2003. Stereoselective hepatic disposition of a diastereomeric pair of alpha v beta3 antagonists in rat. Xenobiotica. 33(11)1125-37.

Put R, Xu QS, Massart DL, Heyden YV. 2004. Multivariate adaptive regression splines (MARS) in chromatographic quantitative structure-retention relationship studies. J Chromatogr A. 1055(1-2)11-19.

Raffa RB. 2001. Drug-receptor thermodynamics: Introduction and application. John Wiley & Sons. pp. 604-20.

Rang HP. 2006. Drug Discovery and Development: Technology in Transition, Churchill Livingstone Elsevier, London.

Rautio J, Humphreys JE, Webster LO, Balakrishnan A, Keogh JP, Kunta JR, Serabjit-Singh CJ, Polli JW. 2006. *In vitro* p-glycoprotein inhibition assays for assessment of clinical drug interaction potential of new drug candidates: a recommendation for probe substrates. Drug Metab Dispos. 34(5)786-92.

RCSB Protein Data Bank, available at: www.rcsb.org/ Last accessed at 21.03.2014

Reid G, Wielinga P, Zelcer N, Van der Heijden I, Kuil A, de Haas M, Wijnholds J, Borst P. 2003. The human multidrug resistance protein MRP4 functions as a prostaglandin efflux transporter and is inhibited by nonsteroidal anti-inflammatory drugs. Proc Natl Acad Sci USA. 100(16)9244-49.

Relling MV. 1996. Are the major effects of P-glycoprotein modulators due to altered pharmacokinetics of anticancer drugs? Ther Drug Monit. 18(4)350-6.

Ren Q, Paulsen IT. 2005. Comparative Analyses of Fundamental Differences in Membrane Transport Capabilities in Prokaryotes and Eukaryotes. PLoS Compt Biol. 1(3)190-201.

Richter O.V, Glavinas H, Krajcsi P, Liehner S, Siewert B, Zech K. 2009. A novel screening strategy to identify ABCB1 substrates and inhibitors. Naunyn Schmiedebergs Arch Pharmacol. 379(1)11-26.

Rius M, Nies AT, Hummel-Eisenbeiss J, Jedlitschky G, Keppler D. 2003. Cotransport of reduced glutathione with bile salts by MRP4 (ABCC4) localized to the basolateral hepatocytes membrane. Hepatol. 38(2)374-84.

Riviere JE. 2011. Comparative pharmacokinetics. Principles, techniques and applications. Second edition. Wiley-Black Well. pp.73-82.

Roberts MS, Magnusson BM, Burczynski FJ, Weiss M. 2002. Enterohepatic Circulation. Physiological, pharmacokinetic and clinical implications. J Clin Pharmacokinetic. 41(10)765-772.

Rognan D. 2013. Proteome-scale docking: myth and reality. Drug Discov Today Technol. 10(3)403-9

Rollins DE, Klaassen CD. 1979. Biliary excretion of drugs in man. J Clin Pharmacokinet. 4(5)368-379.

Rosenbaum SE. 2011. Basic pharmacokinetics and pharmacodynamics, an integrated textbook and computer simulations. 1st edition. John Wiley and Sons. Hoboken, New Jersey.

Rosenberg MF, Callaghan R, Ford RC, Higgins CF. 1997. Structure of the Multidrug Resistance P-glycoprotein to 2.5 nm Resolution Determined by Electron Microscopy and Image Analysis. J Biol Chem. 272(16)10685-94.

Ross DD, Yang W, Abruzzo LV, Dalton WS, Schneider E, Lage H, Dietel M, Greenberger L, Cole SP, Doyle LA. 1999. Atypical multidrug resistance: breast cancer resistance protein messenger RNA expression in mitoxantrone-selected cell lines. J Natl Cancer Inst. 91(5)429-33.

Roth M, Obaidat A, Hagenbuch B. 2012. OATPs, OATs and OCTs: the organic anion and cation transporters of the SLCO and SLC22A gene superfamilies. Br J Pharmacol. 165(5)1260-87.

Rowinsky EK, Cazenave LA, Donehower RC. 1990. Taxon: A novel investigational antimicrotubule agent. J Ntl Cancer Inst. 82(15)1247-59.

Roy K. 2007. On some aspects of validation of predictive quantitative structure-activity relationship models. Expert Opin Drug Discov. 2(12):1567-77.

Saeys Y, Inza I, Larrañaga P. 2007. A review of feature selection techniques in bioinformatics. Bioinformatics. 23(19)2507-17.

Sahigara F, Mansouri K, Ballabio D, Mauri A, Consonni C, Todeschini R. 2012. Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. Molecules. 17(5)4791-4810.

Sai Y, Kaneko Y, Ito S, Mitsuoka K, Kato Y, Tamai I, Artursson P, Tsuji A. 2006. Predominant contribution of organic anion transporting polypeptide OATP-B (OATP2B1) to apical uptake of estrone-3-sulfate by human intestinal Caco-2 cells. Drug Metab Dispos. 34(8)1423-31.

Sainis I, Fokas D, Vareli K, Tzakos AG, Kounnis V, Briasoulis E. 2010. Cyanobacterial cyclopeptides as lead compounds to novel targeted cancer drugs. Mar Drugs. 8(3)629-57.

Saito H, An R, Hirano H, Ishikawa T. 2010. Emerging new technology: QSAR analysis and MO Calculation to characterize interactions of protein kinase inhibitors with the human ABC transporter, ABCG2 (BCRP). Drug Metab Pharmacokinet. 25(1)72-83.

Saito T, Zhang Zj, Tsuzuki H, Ohtsubo T, Yamada T, Yamamato T, Saito H. 1997. Expression of P-glycoprotien in inner ear capillary endothelial cells of the guinea pig with special reference to blood-inner ear barrier. Brain Res. 767(2)388-92.

Sarkadi B, Homolya L, Szakacs G, Varadi A. 2006. Human Multidrug Resistance ABCB and ABCG Transporters: Participation in a Chemoimmunity Defense System. Physiol Rev. 86(4)1179-236.

Sasabe H, Kato Y, Terasaki T, Tsuji A, Sugiyama Y. 1999. Differences in the hepatobiliary transport of two quinolone antibiotics, grepafloxacin and lomefloxacin, in the rat. Biopharm Drug Dispos. 20(3)151-8.

Sata R, Ohtani H, Tsujimoto M, Murakami H, Koyabu N, Nakamura T, Uchiumi T, Kuwano M, Nagata H, Tsukimori K, Nakano H, Sawada Y. 2005. Functional analysis of organic cation transporter 3 expressed in human placenta. J Pharmacol Exp Ther. 315(2)888-95.

Sauer WHB, Schwarz MK. 2003. Molecular Shape Diversity of Combinatorial Libraries: A prerequisite for broad bioactivity. J Chem Inf Comput Sci. 43(3)987-1003.

Schinkel AH, Jonker JW. 2003. Mammalian drug efflux transporters of the ATP binding cassette (ABC) family: an overview. Adv Drug Deliv Rev. 55(1)3-29.

Schinkel AH, Mayer U, Wagenaar E, Mol CA, Deemter LV, Smit JJ, Valk MA, Voordouw AC, Spits H, van Tellingen O, Zijlmans JM, Fibbe WE, Borst P. 1997. Normal viability and altered pharmacokinetics in mice lacking mdr1-type (drug-transporting) P-glycoproteins. Proc Natl Acad Sci USA. 94(8)4028–33.

Schinkel AH, Wagenaar E, van Deemter L. Mol CA, Borst P. 1995. Absence of the mdr 1a P-glycoprotein in mice affects tissue distribution and pharmacokinetics of dexamethasone, digoxin and cyclosporin A. J Clin Invest. 96(4)1698-1705.

Schmidt T, Bergner A, Schwede T. 2013. Modelling three-dimensional protein structures for applications in drug design. Drug Discov Today. 6446(13)1-9.

Schuetz JD, Connelly MC, Sun D, Paibir SG, Flynn PM, Srinivas RV, Kumar A, Fridland A. 1999. MRP4: A previously unidentified factor in resistance to unnucloside-based antiviral drugs. Nat Med. 5(9)1048-51.

Schulz-Gasch T, Stahl M. 2004. Scoring functions for protein–ligand interactions: a critical perspective. Drug Discov Today Tech. 1(3)231-239.

Scott DO, Bindra DS, Sutton SC, Stella VJ. 1994. Urinary and biliary disposition of the lactone and carboxylate forms of 20(S)-camptothecin in rats. Drug Metab Dispos. 22(3)438-42.

Secundo F. 2013. Conformational changes of enzymes upon immobilisation. Chem Soc Rev. 42(15)6250-61.

Sekine T, Cha SH, Endou H. 2000. The multispecific organic anion transporter (OAT) family. Pflugers Arch. 440(3)337-50.

Sekine T, Miyazaki H, Endou H. 2006. Molecular physiology of renal organic anion transporters. Am J Physiol Renal Physiol. 290(2)251-61.

Sekine T, Watanabe N, Hosoyamada M, Kanai Y, Endou H. 1997. Expression cloning and characterization of a novel multispecific organic anion transporter. J Biol Chem. 272(30)18526-9.

Shaik N, Giri N, Pan G, Elmquist WF. 2007. P-glycoprotein-mediated active efflux of the anti-HIV1 nucleoside abacavir limits cellular accumulation and brain distribution. Drug Metab Dispos. 35(11)2076-85.

Shapiro AB, Ling V. 1997. Effect of quercetin on Hoechst 33342 transport by purified and reconstituted P-glycoprotein. Biochem Pharmacol. 53(4)587-96.

Sharifi M, Ghafourian T. 2014. Estimation of biliary excretion of foreign compounds using properties of molecular structure. AAPS J. 16(1)65-78.

Sharom FJ. 2008. ABC multidrug transporters: structure, function and role in chemoresistance. Pharmacogenomics. 9(1)105-27.

Sherlock S, Dooley J. 2008. Diseases of the liver and biliary system. Chapter 1. Anatomy and function. Eleventh edition. Blackwell publication. Oxford, UK. pp. 1-17.

Shin HJ, Anzai N, Enomoto A, He X, Kim do K, Endou H, Kanai Y. 2007. Novel liver-specific organic anion transporter OAT7 that operates the exchange of sulfate conjugates for short chain fatty acid butyrate. Hepatology. 45(4)1046-55.

Shirasaka Y, Mori T, Murata Y, Nakanishi T, Tamai I. 2014. Substrate- and Dose-Dependent Drug Interactions with Grapefruit Juice Caused by Multiple Binding Sites on OATP2B1. Pharm Res. 2014 Feb 19. [Epub ahead of print].

Shitara Y, Itoh T, Sato H, Li AP, Sugiyama Y. 2003. Inhibition of transporter-mediated hepatic uptake as a mechanism for drug-drug interaction between cerivastatin and cyclosporin A. J Pharmacol Exp Ther. 304(2)610-6.

Shitara Y, Maeda K, Ikejiri K, Yoshida K, Horie T, Sugiyama Y. 2013. Clinical significance of organic anion transporting polypeptides (OATPs) in drug disposition: their roles in hepatic clearance and intestinal absorption. Biopharm Drug Dispos. 34(1)45-78.

Shitara Y, Sugiyama D, Kusuhara H, Kato Y, Abe T, Meier PJ, Itoh T, Sugiyama Y. 2002. Comparative inhibitory effects of different compounds on rat oatpl (slc21a1)- and Oatp2 (Slc21a5)-mediated transport. Pharm Res. 19(2)147-53.

Siegel R, Naishadham D, Jemal A. 2013. Cancer statistics, 2013. CA Cancer J Clin. 63(1)11-30.

Smith JH. 2006. Introduction to the principles of drug design and action. In: Gumbleton M. editor, Processes of drug handling by the body. Third edition. Harwood academic publishers. FL, USA. pp. 24-25.

Smith PA, Sorich MJ, McKinnon RA, Miners JO. 2003. *In silico* insights: chemical and structural characteristics associated with uridine diphosphate glucuronosyltransferase substrate selectivity. Clin Exp Pharmacol Physiol. 30(11)836-40.

Soars MG, Barton P, Ismair M, Jupp R, Riley RJ. 2012. The development, characterization, and application of an OATP1B1 inhibition assay in drug discovery. Drug Metab Dispos. 40(8)1641-8.

Song B, Wang Y, Titmus MA, Botchkina G, Formentini A, Kornmann M, Ju J. 2010. Molecular mechanism of chemoresistance by miR-215 in osteosarcoma and colon cancer cells. Mol Cancer. 9(96)1-10.

Song S, Suzuki H, Kawai R, Tanaka C, Akasaka I, Sugiyama Y. 1998. Dose-dependent effects of PSC 833 on its tissue distribution and on the biliary excretion of endogenous substrates in rats. Drug Metab Dispos. 26(11)1128-33.

Sousa SF, Fernandes PA, Ramos MJ. 2006. Protein–Ligand Docking: Current Status and Future Challenges. Proteins. 65(1)15-26.

Stanton DT. 1999. Evaluation and Use of BCUT Descriptors in QSAR and QSPR Studies. J Chem Inf Comput Sci. 39(1)11-20.

StatSoft. Elecctronic Statistics Textbook. 2009. Available from: http://www.statsoft.com/textbook/chaid-analysis/?button=1 (Accessed on 06 March 2014).

Stewart JJP. MOPAC Manual: A General Molecular Orbital Package. Seventh Edition. New York. 1993.

Stouch TR, Kenyon JR, Johnson SR, Chen XQ, Doweyko A, Li Y. 2003. *In silico* ADME/Tox: why models fail. J Comput Aided Mol Des. 17(2-4)83-92.

Stringer JL. Basic concepts in pharmacology: a student's survival guide. 3-rd edition, The Mc.Graw-hill companies. 2006. Chapters 3 and 5. pp. 18-30.

Sugawara M, Huang W, Fei YJ, Leibach FH, Ganapathy V, Ganapathy ME. 2000. Transport of valganciclovir, a ganciclovir prodrug, via peptide transporters PEPT1 and PEPT2. J Pharm Sci. 89(6)781-9.

Sun H, Huang Y, Frassetto L, Benet LZ. 2004. Effects of uremic toxins on hepatic uptake and metabolism of erythromycin. Drug Metab Dispos. 32(11) 1239-46.

Sun W, Wu RR, van Poelje PD, Erion MD. 2001. Isolation of a family of organic anion transporters from human liver and kidney. Biochem Biophys Res Commun. 283(2)417-22.

Svoboda M, Wlcek K, Taferner B, Hering S, Stieger B, Tong D, Zeillinger R, Thalhammer T, Jäger W. 2011. Expression of organic anion-transporting polypeptides 1B1 and 1B3 in ovarian cancer cells: relevance for paclitaxel transport. Biomed Pharmacother. 65(6)417-26.

Taft D. Drug excretion. In: Hacker M, Bachmann K, Messer W (editors). Pharmacology, principles and practice. Oxford: Elsevier Inc; 2009. pp. 175-199.

Takekuma Y, Kakiuchi H, Yamazaki K, Miyauchi S, Kikukawa T, Kamo N, Ganapathy V, Sugawara M. 2007. Difference between pharmacokinetics of mycophenolic acid (MPA) in rats and that in humans is caused by different affinities of MRP2 to a glucuronized form. J Pharm Pharm Sci. 10(1)71-85.

Tamai I, Yabuuchi H, Nezu J, Sai Y, Oku A, Shimane M, Tsuji A. 1998. Molecular and functional identification of sodium ion-dependent, high affinity human carnetine transporter OCTN2. J Biol Chem. 273(32)20378-82.

Tang F, Horie K, Borchardt RT. 2002a. Are MDCK cells transfected with the human MRP2 gene a good model of the human intestinal mucosa? Pharm Res. 19(6)773-79.

Tang F, Horie K, Borchardt RT. 2002b. Are MDCK cells transfected with the human MDR1 gene a good model of the human intestinal mucosa? Pharm Res. 19(6)765-72.

Taub ME, Mease K, Sane RS, Watson CA, Chen L, Ellens H, Hirakawa B, Reyner EL, Jani M, Lee CA. 2011. Digoxin is not a substrate for organic anion-transporting polypeptide transporters OATP1A2, OATP1B1, OATP1B3, and OATP2B1 but is a substrate for a sodium-dependent transporter expressed in HEK293 cells. Drug Metab Dispos. 39(11)2093-102.

Teague SJ. 2003. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov. 2(7)527-541.

Terada T, Masuda S, Asaka J, Tsuda M, Katsura T, Inui K. 2006. Molecular cloning, functional characterization and tissue distribution of rat H^+/organic cation antiporters MATE1. J Pharm Res. 23(8)1696-1701.

Thakur A, Thakur M, Khadikar PV, Supuran CT, Sudele P. 2004. QSAR study on benzenesulphonamide carbonic anhydrase inhibitors: topological approach using Balaban index. Bioorg Med Chem. 12(4)789-93.

Thiebaut F, Tsuruo T, Hamada H, Gottesman MM, Pastan I, Willingham MC. 1987. Cellular localization of the multidrug-resistance gene product P-glycoprotein in normal human tissues. Proc Natl Acad Sci USA. 84(21)7735-38.

Tirona RG, Leake BF, Merino G, Kim RB. 2001. Polymorphisms in OATP-C: identification of multiple allelic variants associated with altered transport activity among European- and African-Americans. J Biol Chem. 276(38)35669-75.

Todeschini R, Consonni V. Handbook of molecular descriptors. 2008. Molecular descriptors. Germany: Volume 11. Wiley-VCH; 2008. pp. 303.

Toyoda Y, Hagiya Y, Adachi T, Hoshijima K, Kuo MT, Ishikawa T. 2008. MRP class of human ATP binding cassette (ABC) transporters: historical background and new research directions. Xenobiotica. 38(7-8)833-62.

Tozer TN, Rowland M. Introduction to pharmacokinetics and pharmacodynamics. 2006. Lippincott Williams and Wilkins, NewYork.

Trauner M, Arrese M, Soroka CJ, Ananthanarayanan M, Koeppel TA, Schlosser SF, Suchy FJ, Keppler D, Boyer JL. 1997. The rat canalicular conjugate export pump (Mrp2) is down-regulated in intrahepatic and obstructive cholestasis. Gastroenterology. 113(1)255-64.

Trauner M, Boyer JL. 2003. Bile Salt Transporters: Molecular characterization, function, and regulation. Physiol Rev. 83(2)633-71.

Tropsha A, Zheng W. 2001. Identification of the descriptor pharmacophores using variable selection QSAR: applications to database mining. Curr Pharm Des. 7(7)599-612.

Tropsha, A, Cho S, Zheng W. In: Rational Drug Design: Novel Methodology and Practical Applications (Parrill AL, Reddy MR. Eds), ACS Symposium Series. No 719, 1999, pp. 198-211.

Tsaioun K, Kates SA. ADMET for medicinal chemists. A practical guide. First edition. John Wiley & Sons. 2011. Hoboken, New Jersey.

Urbatsch IL, Sankaran B, Bhagat S, Senior AE. 1995. Both P-glycoprotein nucleotide-binding sites are catalytically active. J Biol Chem. 270(45)26956-61.

Vaidyanathan S, Boroujerdi M. 2000. Effect of tamoxifen pretreatment on the pharmacokinetics, metabolism and cardiotoxicity of doxorubicin in female rats. Cancer Chemother Pharmacol. 46(3)185-92.

van Aubel RA, Smeets PH, Peters JG, Bindels RJ, Russel FG. 2002. The MRP4/ABCC4 gene encodes a novel apical organic anion transporter in human kidney proximal tubules: putative efflux pump for urinary cAMP and cGMP. J Am Soc Nephrol. 13(3)595-603.

Van de Waterbeemd H, Rose S. 2003. The Practice of Medicinal Chemistry Second edition. Oxford: Academic Press.

van de Waterbeemd H, Smith DA, Jones BC. 2001. Lipophilicity in PK design: methyl, ethyl, futile. J Comput Aided Mol Des. 15(3)273-86.

van Herwaarden AE, Schinkel AH. 2006. The function of breast cancer resistance protein in epithelial barriers, stem cells and milk secretion of drugs and xenotoxins. Trends Pharmacol Sci. 27(1)10-6.

van Montfoort JE, Hagenbuch B, Groothuis GM, Koepsell H, Meier PJ, Meijer DK. 2003. Drug uptake systems in liver and kidney. Curr Drug Metab. 4(3)185-211.

Varma MV, Ashokraj Y, Dey CS, Panchagnula R. 2003. P-glycoprotein inhibitors and their screening: a perspective from bioavailability enhancement. Pharmacol Res. 48(4)347-59.

Varma MV, Chang G, Lai Y, Feng B, El-Kattan AF, Litchfield J, Goosen TC. 2012. Physicochemical property space of hepatobiliary transport and computational models for predicting rat biliary excretion. Drug Metab Dispos. 40(8)1527-37.

Varma MV, Rotter CJ, Chupka J, Whalen KM, Duignan DB, Feng B, Litchfield J, Goosen TC, El-Kattan AF. 2011. pH-sensitive interaction of HMG-CoA reductase inhibitors (statins) with organic anion transporting polypeptide 2B1. Mol Pharm. 8(4)1303-13.

Vereczkey L, Szporny L. 1980. Disposition of pipecurium bromide in rats. Arzneimittel-forschung. 30(2)364-6.

Verma RP, Hansch C. 2005. A Comparison between two polarisability parameters in chemical–biological interactions. Bioorg Med Chem. 13(7)2355–72.

Wacher VJ, Silverman JA, Zhang Y, Benet LZ. 1998. Role of P-glycoprotein and cytochrome P450 3A in limiting oral absorption of peptides and peptidomimetics. J Pharm Sci. 87(11)1322–30.

Wacher VJ, Wu CY, Benet LZ. 1995. Overlapping substrate specificities and tissue distribution of Cytochrome P450 3A and P-glycoprotein: Implications for drug delivery and activity in cancer chemotherapy. Mol Carcinog. 13(3)129-34.

Wada S, Tsuda M, Sekine T, Cha SH, Kimura M, Kanai Y, Endou H. 2000. Rat multispecific organic anion transporter 1 (rOAT1) transports zidovudine, acyclovir, and other antiviral nucleoside analogs. J Pharmacol Exp Ther. 294(3)844-9.

Wandel C, Kim RB, Kajiji S, Guengerich FP, Wilkinson GR, Wood AJ. 1999. P-Glycoprotein and cytochrome P-450 3A inhibition: Dissociation of inhibitory potencies. Cancer Res. 59(16)3944-48.

Wang E, Casciano CN, Clement RP, Johnson WW. 2001. The farnesyl protein transferase inhibitor SCH66336 is a potent inhibitor of MDR1 product P-glycoprotein. Cancer Res. 61(20)7525-9.

Wang RB, Kuo CL, Lien LL, Lien EJ. 2003. Structure-activity relationship: analyses of p-glycoprotein substrates and inhibitors. J Clin Pharm Ther. 28(3)203-28.

Wang Z, Chen Y, Liang H, Bender A, Glen RC, Yan A. 2011. P-glycoprotein substrate models using support vector machines based on a comprehensive dataset. J Chem Inf Model. 51(6)1447-56.

Watkins JB, Dykstra TP. 1978. Alterations in biliary excretory function by streptozotocin-induced diabetes. Drug Metab Dispos. 15(2)177–183.

Weaver DC. 2004. Applying data mining techniques to library design, lead generation and lead optimization. Curr Opin Chem Biol. 8(3)264-70.

Weaver S, Gleeson MP. 2008. The importance of the domain of applicability in QSAR modeling. J Mol Graph Model. 26(8)1315-26.

Weinz C, Schwarz T, Kubitza D, Mueck W, Lang D. 2009. Metabolism and Excretion of Rivaroxaban, an Oral, Direct Factor Xa Inhibitor, in Rats, Dogs, and Humans. Drug Metab Dispos. 37(5)1056-64.

Williams RT. 1959. Detoxication mechanisms. Second edition. New York. John Wiley & Sons.

Wold S, Eriksson L. 1995. Statistical validation of QSAR results. In Waterbeemd Han. Chemometric methods in molecular design. Weinheim: VCH. pp. 309–318.

Wong CM, Ko Y, Chan A. 2008. Clinically significant drug-drug interactions between oral anticancer agents and nonanticancer agents: profiling and comparison of two drug compendia. Ann Pharmacother. 42(12)1737-48.

Wright EW, Line VD. 1980. Biliary excretion of cephalosporins in rats: Influence of molecular weight. Antimicrob Agents Chemother. 17(5)842-46.

Wu WM, Huang F, Lee Y, Buchwald P, Bodor N. 2008. Pharmacokinetics of the sequential metabolites of loteprednol etabonate in rats. J Pharm Pharmacol. 60(3)291-97.

Wu X, George RL, Huang W, Wang H, Conway SJ, Leibach FH, Ganapathy V. 2000. Structural and functional characteristics and tissue distribution pattern of rat OCTN1, an organic cation transporter, cloned from placenta. Biochim Biophys Acta. 1466(1-2)315-327.

Wu X, Wei H, Puttur D, Puttur DP, Pankaj S, Deva PR, Frederick HL, Jinwen C, Simon JC, Vadivel G. 1999. Functional characteristics and tissue distribution pattern of organic cation transporter 2 (OCTN2), an organic cation/carnitine transporter. J Pharmacol Exp Ther. 290(3)1482-92.

Xu Y, Gao H. 2003. Dimension related distance and its application in QSAR/QSPR model error estimation. QSAR Comb Sci. 22(4)422-429.

Yabuuchi H, Shimizu H, Takayanagi S, Ishikawa T. 2001. Multiple splicing variants of two new human ATP-binding cassette transporters, ABCC11 and ABCC12. Biochem Biophys Res Commun. 288(4)933-9.

Yamazaki M, Suzuki H, Sugiyama Y. 1996. Recent advances in carrier-mediated hepatic uptake and biliary excretion of xenobiotics. Pharm Res. 13(4)497-513.

Yan Z, Caldwell GW. 2001. Metabolism profiling, and cytochrome P450 inhibition & induction in drug discovery. Curr Top Med Chem. 1(5)403-25.

Yang X, Gandhi YA, Duignan DB, Morris ME. 2009. Prediction of Biliary Excretion in Rats and Humans Using Molecular Weight and Quantitative Structure–Pharmacokinetic Relationships. AAPS J. 11(3)511-25.

Yang X, Gandhi YA, Morris ME. 2010. Biliary excretion in dogs: Evidence for a molecular weight threshold. Europ J Pharm Sci. 40(1)33-37.

You G, Morris ME. 2007. Drug transporter, molecular characterization and role in drug disposition. John Wiley publication. First edition. Hoboken, New Jersey.

Yu M, Zhang WG, Qin LH, Tian L, Zhou, CM. 2010. Enhancement of P-Glycoprotein Expression by Hepatocyte Transplantation in Carbon Tetrachloride-Induced Rat Liver. Anat Rec (Hoboken). 293(7)1167-74.

Zamek-Gliszczynski MJ, Hoffmaster KA, Humphreys JE, Tian X, Nezasa K, Brouwer KLR. 2006. Differential Involvement of Mrp2 (Abcc2) and Bcrp (Abcg2) in Biliary Excretion of 4-Methylumbelliferyl Glucuronide and Sulfate in the Rat. J Pharmacol Exp Ther. 319(1)459-67.

Zhang L, Balimane PV, Johnson SR, Chong S. 2007. Development of an *in silico* model for predicting efflux substrates in Caco-2 cells. Int J Pharma. 43(1-2)98-105.

10. Appendix

Appendix I. Percentage of compound's dose excreted intact through the bile in rats and the relevant references

Compounds	BE%	Reference
1,2,3,6-Tetrahydrophthalylsulphathiazole	45.00	Hirom PC, et.al. Biochem J. 1972 Oct; 129(5):1071-7. Female Wistar albino rats (180-350g body wt.)
17-AAG(NSC 330507)	2.00	Musser SM, et.al. Cancer Chemother Pharmacol. 2003 Aug;52(2):139-46. Male Fischer 344 rats (7-8 weeks of age and weighing 220-234g)
17-DMAG (NSC 707545)	2.38	Egorin MJ, et.al. Cancer Chemother Pharmacol. 2002 Jan;49(1):7-19. Male Fischer 344 rats (7-8 weeks of age). % of Dose (in total): 4.7 ± 1.4. Parent drug accounted for 50.7 ± 3.4% of that.
2-Aminotoluene-5-sulphonic acid	0.27	McMahon KA, et.al. Food Cosmet Toxicol. 1969 Sep;7(5):497-500. Rats (250-350 g body weight)
2-Ethylsulphanilic acid	0.29	McMahon KA, et.al. Food Cosmet Toxicol. 1969. Sep;7(5):497-500. Rats (250-350 g body weight)
4-Glucuronosido-4'-hydroxybiphenyl	92.00	Millburn P., et al, Biochem. J. 1967; 105, 1275 Female Wistar albino rats (weighing 200 ± 10g.)
4-Glucuronosidobiphenyl	59.00	Millburn P., et al, Biochem. J. 1967; 105, 1275 Female Wistar albino rats (weighing 200 ± 10g.)
5-fluorouracil (5-FU)	0.40	Young D, et.al. Nuklearmedizin. 1982 Feb;21(1):1-7. Male Fischer rats weighing 150 - 200g
7-Hydroxymethotrexate	37.00	Lutz Fahrig, Helmut Brasch, et al, Cancer Chemother Pharmacol(1989)23, 156-160
9-nitro-20(S)-camptothecin(Rubitecan)	9.10	Zhong DF, et.al. Acta Pharmacol Sin. 2003 Mar;24(3):256-62. Wistar rats (250 ± 20g)
acetaminophen(paracetamol)	0.80	Ghanem CI, et.al. J Pharmacol Exp Ther. 2005 Dec;315(3):987-95. Male Wistar rats (250-290 g) Savina PM, et.al. Drug Metab Dispos. 1992 Jul-Aug;20(4):496-501. Male Sprague-Dawley rats (266-282 g).
actinomycin D	31.00	Wosilait WD, et.al. Life Sci I. 1971 Sep 15;10(18):1051-5 Male Sprague-Dawley rats, weighing about 300 g.
adipylsulphathiazole	40.00	Hirom PC, et.al. Biochem J. 1972 Oct;129(5):1071-7. Female Wistar albino rats (180-350g body wt.)
aprepitant	7.00	Huskey SE, et.al. Drug Metab Dispos. 2004 Feb;32(2):246-58. Male SD rats (230-300 g)
azithromycin	9.60	Sugie M, et.al. Antimicrob Agents Chemother. 2004 Mar;48(3):809-14. Male Wistar Rats, 260 - 270g. Male Sprague-Dawley rats (normal rats) (260 to 280g)
belotecan	28.29	Namkoong EM, et.al. Arch Pharm Res. 2007 Nov;30(11):1482-8. Male SD rats (260 - 290g)
Benzoic acid	0.09	Abou-el-makarem M.M., Millburn P, et al Biochem. J.(1967)105, 1269
beta-methyldigoxin	53.00	Funakoshi S., Murakami T, et al, J Pharm Sci. (2005)94(6), 1196-203
bishydroxycoumarin	1.88	Buttar HS, et.al. Br J Pharmacol. 1973 Jun;48(2):278-87. Male Albino rats (Wistar, 275 - 355g). % of Dose (in total): 12.3 ± 2.7, Parent drug accounted for 15.3

Compounds	BE%	Reference
		(12.5-18.4) % of that.
BMS-182874	0.90	Chong s, Obermeier M et al. 2003. Arch pharm sci 26:89-94.
BMS-187345	4.50	Chong s, Obermeier M et al. 2003. Arch pharm sci 26:89-94.
BMS-387032	11.00	Kamath AV chong S et al. 2005. Cancer chemother pharmacol 55:110-116.
BQ-123	52.82	Kato Y, et.al. J Pharmacol Exp Ther. 1999 Feb;288(2):568-74. Male Sprague-Dawley rats weighing approximately 250 to 300g . Nakamura T, et.al. J Pharmacol Exp Ther. 1996 Aug;278(2):564-72. Male Sprague-Dawley rats, 7 to 10 weeks of age. Niinuma K, Kato Y, et al. Am J Physiol. 1999 ;276(5 Pt 1)1153-1164.
BQ-485	97.40	Kato Y, et.al. J Pharmacol Exp Ther. 1999 Feb;288(2):568-74. Male Sprague-Dawley rats weighing approximately 250 to 300g
BQ-518	89.70	Kato Y, et.al. J Pharmacol Exp Ther. 1999 Feb;288(2):568-74. Male Sprague-Dawley rats weighing approximately 250 to 300g
bretylium	16.00	Kuntzman R, et.al. Clin Pharmacol Ther. 1970 Nov-Dec;11(6):829-37
bromochlorophenol blue	89.00	Hirom PC, et.al. Biochem J. 1972 Oct;129(5):1071-7. Female Wistar albino rats (180-350g body wt.)
Bromocresol Green	73.00	Hirom PC, et.al. Biochem J. 1972 Oct;129(5):1071-7. Female Wistar albino rats (180-350g body wt.)
Bromophenol Blue	67.25	Hirom PC, et.al. Biochem J. 1972 Oct;129(5):1071-7. Female Wistar albino rats (180-350g body wt.) Wills RJ, et.al. J Pharm Sci. 1983 Oct;72(10):1127-31 Fasted male Sprague-Dawley rats (260 - 470g)
buprenorphine	1.08	Brewster D, et.al. Xenobiotica. 1981 Mar;11(3):189-96. Adult SD rats (200-300g). % of Dose (in total): 92.9 + 8.0. Parent drug accounted for 1.5 ± 0.8% of that (Male) % of Dose (in total): 94.5 ± 2.8, Parent drug accounted for 0.8 ± 0.4% of that. (female).
butoprozine	0.00	Overzet F, et.al. Xenobiotica. 1985 Jan;15(1):1-10. male Wistar rats (body wt. 300g)
cadrala zine	3.70	Eur J Drug Metab Pharmacokinet. 1983;8(1):25-33. Male and female Sprague Dawley rats with an average body weight of 150 to 180 g.
camptothecin (carboxylate form)	36.40	Scott DO, et.al. Drug Metab Dispos. 1994 May-Jun;22(3):438-42. Male Sprague-Dawley rats weighing between 250-300g. Guarino AM, et.al. Cancer Chemother Rep. 1973 Apr;57(2):125-40. Male Sprague-Dawley rats (240 - 320g)
camptothecin (lactone form)	7.50	Scott DO, et.al. Drug Metab Dispos. 1994 May-Jun;22(3):438-42. Male Sprague-Dawley rats weighing between 250-300g.
carbovir	1.30	Zimmerman CL, et.al. Drug Metab Dispos. 1993 Sep-Oct;21(5):902-10. Sprague-Dawley rat
cefamandole	33.00	Wright WE, et.al. Antimicrob Agents Chemother. 1980 May;17(5):842-6. Male Wistar rats, weighing 350 to 500g
cefazedone	37.40	Sailer H, et.al. Arzneimittelforschung. 1979;29(2a):404-11

Compounds	BE%	Reference
		Male and female Wistar-WU rats (weight range 175-320g.
cefazolin	30.00	Tsuji A, et.al. J Pharm Sci. 1983 Nov;72(11):1239-52. Male Wistar rats (240g)
cefbuperazone (T-1982)	80.00	Saikawa I, et.al. Jpn J Antibiot. 1982 Sep;35(9):2163-73
cefixime	40.80	Yasui H, et.al. J Pharm Sci. 1994 Jun;83(6):819-23 Male Witar rats (177 - 230g).
cefmenoxime (SCE-1365)	28.50	Tanayama S, et.al. Antimicrob Agents Chemother. 1980 Oct;18(4):511-8. male or female Sprague-Dawley rats weighing 220 to 515g.
cefmetazole	36.25	Eur J Drug Metab Pharmacinet. 1992 Jul-Sep;17(3):167-73. Male Wistar: 232-298g.
cefodizime	28.60	Matsushita H, et.al. J Pharmacol Exp Ther. 1992 Feb;260(2):499-504. Male Wistar rats weighing 240 to 280g.
cefoperazone	85.60	Saikawa I, et.al. Jpn J Antibiot. 1980 Oct;33(10):1084-96
cefotetan (YM-09330)	48.00	Komiya M, et.al. Antimicrob Agents Chemother. 1981 Aug;20(2):176-83. SD rats: 200 - 350g. Mizojiri K, et.al. Antimicrob Agents Chemother. 1987 Aug;31(8):1169-76
cefpiramide (SM-1652)	59.80	Matsui H, et.al. Antimicrob Agents Chemother. 1982 Aug;22(2):213-7. Male Sprague-Dawley rats (200 to 250 g). Imasaki H, et.al. Antimicrob Agents Chemother. 1983 Jul;24(1):42-7. Sprague-Dawley male rats weighting 150 to 300g. Muraoka I, et.al. Antimicrob Agents Chemother. 1995 Jan;39(1):70-4. 20-week-old healthy SDR (weight, 494 to 540 g)
ceftriaxone	61.80	Matsui H, et.al. Antimicrob Agents Chemother. 1984 Aug;26(2):204-7. male SD rats (body weight, 200 to 250 g)
celiptium (NSC-264137)	6.10	Maftouh M, et.al. Xenobiotica. 1983 May;13(5):303-10 Male SD rats (300 - 350g).
Cephalexin	2.50	Wright W.E., Line V.D. Antimicrobial Agents & Chemotherapy(1980)17, 842-846. Male Wistar rats [HAP(WI)BR], weighing 350 to 500
Cephradine	27.30	Moriwaki T, Yasui H and Yamamoto A. 2003. J Pharmacokinet Phamacodyn 30:119-144.
chenodeoxycholate (CDC)	0.30	Takikawa H,et.al. Hepatology. 1991. 14(2):352-60. Male SDRs weighing about 270g. % of dose (in total): ~ 3% at steady state. Parent drug accounted for 6 -10% of that.
ciprofloxacin	9.92	Yamaguchi H, et.al. Pharm Res. 2004 Feb;21(2):330-8. Male Wistar rats weighing 200-250g
colchicine	25.36	Hunter AL, et.al. J Pharmacol Exp Ther. 1975 Mar;192(3):605-17. Male Thorp SD rats (350-390g). % of dose (in total): 52. Parent drug accounted for 53% of that. Speeg KV, et.al. Cancer Chemother Pharmacol. 1994;34(2):133-6. Male Sprague-Dawley rats weighing 300-400g . Speeg KV, et.al. Hepatology. 1992 May;15(5):899-903. Male SD rats weighing 300 to 400g. CLsys: 43.05 ± 2.68 ml/min/kg. CLbiliary: 11.62 ± 0.84 ml/min/ Kitani K, et.al. Tohu J Exp Med. 1981 Apr;133(4):389-

Compounds	BE%	Reference
		97. Male Wistar rats (300g on the average). % of dose (in total): 35.19 ± 2.91. Parent drug accounted for 70.82 ± 7.79% of that.
compound I (Merck) diastereomer	13.00	Prueksaritanont T, et.al. Xenobiotica. 2003 Nov;33(11):1125-37. Male Sprague–Dawley (SD) rats (200–320g). Prueksaritanont T, et.al. Xenobioticaxenobiotica,2002, vol. 32, no. 3, 207±220 Male Sprague–Dawley (SD) rats (230–320g).
compound II (Merck) diastereomer	58.00	Prueksaritanont T, et.al. Xenobiotica. 2003 Nov;33(11):1125-37. Male Sprague–Dawley (SD) rats (200–320g). Prueksaritanont T, et.al. Xenobioticaxenobiotica,2002, vol. 32, no. 3, 207±220 Male Sprague–Dawley (SD) rats (230–320g).
cosalane	1.12	Kuchimanchi KR, et.al. Drug Metab Dispos. 2000 Apr;28(4):403-8. Male SD rats weighing 200 to 225 g
CP-671,305	48.33	Kalgutkar AS, et.al. Xenobiotica. 2004 Aug;34(8):755-70 Male and female Sprague-Dawley rats (220-250g) Kalgutkar AS, et.al. Drug Metab Dispos. 2007 35(11):2111-8. Male Sprague-Dawley rats (230-250g)
cromoglycate	71.40	Ashton MJ, et.al. Toxicol Appl Pharmacol. 1973 Nov;26(3):319-28 Male Sprague-Dawley rats (200 - 250g)
DA-5018 (Capsavanil)	3.06	Shim HJ, et.al. J Chromatogr B Biomed Sci Appl. 1997 Feb 21;689(2):422-6.
dasatinib	10.40	Christopher LJ, et.al. Drug Metab Dispos. 2008 Jul;36(7):1341-56. male Sprague-Dawley rats weighing approximately 340 to 380g. G.Luo, S.Johnson, et al, Drug Metab Dispos. J. (2010)38, 422-430
daunorubicin	11.76	Yesair DW, et.al. Cancer Res. 1972 Jun;32(6):1177-83 Male Sprague-Dawley rats (350 to 500 g). Amount excreted into bile: ~ 500µg. Dose: 10 mg/kg.
Decamethonium bromide	1.00	Hughes R.D., Millburn P., et al, Biochem. J. (1973)136, 979-984
diazepam	0.00	Inaba T, et.al. Drug Metab Dispos. 1974 Sep-Oct;2(5):429-32. Male Wistar rats (280-320 g).% of Dose (in total): 77; No intact diazepam could be detected in bile.
Dibenzyldimethylammonium iodide	36.00	Hughes RD, Millburn P. et al, Biochem. J. (1973)136, 967-78
diclofenac	2.99	Peris-Ribera JE, et.al. J Pharmacinet Biopharm. 1991 Dec;19(6):647-65. Male Wistar rats (320-380 g). % of Dose (in total): 27.2; Parent drug accounted for 4.7% of that.
Diethylmethylphenylammonium iodide	7.60	Hughes RD, Millburn P. et al, Biochem. J. (1973)136, 967-78
digoxin	84.4	Song S, et.al. Drug Metab Dispos. 1999 Jun;27(6):689-94 Female Sprague-Dawley (SD) rats weighing 220 to 270g. S. Funakoshi, T. Murakami, et al, J Pharm Sci. (2005)94(6), 1196-203 H.Fukuda, R.Ohashi, et al,Drug Metab Dispos. 2008 Jul;36(7):1275-82
Dimethyltubocurarine iodide	17.00	Hughes RD., Millburn P, et al, Biochem. J. (1973)136,

Compounds	BE%	Reference
		979-984.
DNP-NAC	42.00	Hinchman CA., Rebbeor JF et. 1998. Am j physiol 275(4 pt 1): G612-9.
DNP-SG(2,4-Dinitrophenyl-S-glutathione)	100	Niinuma K, Kato Y, et al American journal of physiology: Gastrointestinal & liver physiology, 1999 ;276(5 Pt 1)1153-1164.
doxorubicin	18.26	Vaidyanathan S, et.al. Cancer Chemother Pharmacol. 2000;46(3):185-92. Female Sprague-Dawley rats weighing 225 to 250g. Krishna R, et.al. Clin Cancer Res. 1999 Oct;5(10):2939-47. Male SD rats, 225-275g. Broggini M, et.al. Cancer Treat Rep. 1980. 64(8-9):897-904. CD-COBS male rats (body weight, 200 ± 20 g) Israel M, et.al. Cancer Res. 1978 Feb;38(2):365-70. Male SD rats weighing 320 to 440 g. % of Dose (in total): 20; Parent drug accounted for 80% of that.
DPDPE	80.00	Chen C, et.al. Pharm Res. 1997 Mar;14(3):345-50 Male Sprague-Dawley rats (250-300g
drotaverine	0.00	Vargay Z., Simon G., et al. Eur J Drug Metab Pharmacokinet. 1980;5(2):69-74
E3040 glucuronide	90.00	Niinuma K., Kato Y, et al American journal of physiology: Gastrointestinal & liver physiology, 1999 ;276(5 Pt 1)1153-1164. Takenaka O, Horie T, Suzuki H, Sugiyama Y, J Pharmacol Exper Ther. 280(2), 948-958. Male SD rats (250–330 g) from Japan Laboratory Animals Inc. Hirouchi M et al, Drug Metab Disp. 37 (10)2103-2111; OCT 2009, Male Mrp3(- /-) mice and wild-type FVB mice (12–18 weeks).
edatrexate	43.35	Fanucchi MP, et.al. Cancer Res. 1987 May 1;47(9):2334-9 Male CD rats. % of Dose (in total): 51 ± 4; Parent drug accounted for 85% of that.
EDDP	36.00	Baselt RC, et.al. Biochem Pharmacol. 1973 Dec 1;22(23):3117-20. Sprague-Dawley male rats (200 - 300 g).
EMDP	0.20	Baselt RC, et.al. Biochem Pharmacol. 1973 Dec 1;22(23):3117-20. Sprague-Dawley male rats (200 - 300 g).
emepronium (EME)	12.00	Neef C, et.al. Naunyn Schmiedebergs Arch Pharmacol. 1984 Dec;328(2):103-10. Male Wistar rats (approximately 300g). % of Dose (in total): 60; Parent drug accounted for < 20% of that
epirubicin (4'-epiDOX)	20.00	Broggini M, et.al. Cancer Treat Rep. 1980 Aug-Sep;64(8-9):897-904. CD-COBS male rats (body weight, 200 ± 20 g)
erythromycin	32.20	Akashi M, et.al. Hepatol Res. 2006 Feb 11,193-198 Male Sprague-Dawley rats weighting approximately 270g. Kageyama M, et.al. Biol Pharm Bull. 2005 Feb;28(2):316-22. Male Wistar rats (280 to 320g). Amount excreted into bile: 200.3 ± 35.6 µg. Dose: 3 mg/kg. Sato A, et.al. Pharmacology. 1999 Nov;59(5):249-56 Tachizawa H, et.al. J Gastroenterol Hepatol. 2004 Sep;19(9):1016-22. Male Sprague-Dawley rats 270g. Lam JL, et.al. Drug Metab Dispos. 2006

Compounds	BE%	Reference
		Aug;34(8):1336-44. Male Wistar rats (200 - 350g). CLtot: 47.2 ± 12.5 and 42.1 ± 5.7 ml/min/kg. CLbiliary: 15.5 ± 2.9 and 11.2 ± 2.0 ml/min/kg.
Estradiol-17ß-glucuronide	87.00	Akashi M, et.al. Hepatol Res. 2006 Feb 11,193-198
Estrone 3-sulphate	18.40	H.Fukuda, R.Ohashi, et al,Drug Metab Dispos. 2008 Jul;36(7):1275-82. Male Sprague-Dawley rats (Charles River Japan, Yokohama. Japan) weighing 200 to 250 g
felodipine	0.00	Sutfin TA, et.al. Xenobiotica. 1987 Oct;17(10):1203-14. Male SD rats (350g). % of Dose (in total): 74; No unchanged felodipine was detected in either bile.
fexofenadine	55.05	Tahara H, et.al. Drug Metab Dispos. 2005 Jul;33(7):963-8 SD rats, 300-350g. CLtot: 28.3 ± 2.1 ml/min/kg; $CL_{biliary}$: 11.4 ± 1.6 ml/min/kg Tian X., Swift B. Drug Metab Dispos. (2008)36(5), 911-915
floctafenin	8.90	Pottier J, et.al. Drug Metab Dispos. 1975 May-Jun;3(3):133-47. Wistar of Sprague-Dawley rats (200 g).
flomoxef	17.50	Hishikawa S, et.al. Chronobiol Int. 2003 May;20(3):463-71. Male Wistar rats weighing 250-300 g
fluvastatin	19.50	Lindahl A, et.al. Mol Pharm. 2004 Sep-Oct;1(5):347-56 Male Sprague-Dawley rats (305 ± 20g
fosmidomycin	0.10	Murakawa T, et.al. Antimicrob Agents Chemother. 1982 Feb;21(2):224-30.
FPL 55712	50.00	Mead B, et.al. J Pharm Pharmacol. 1981 Oct;33(10):682-4 Male Wistar rats
furosemide	1.17	Chen C, et.al. Pharm Res. 2003 Jan;20(1):31-7. Male Sprague-Dawley rats, 15 weeks of age (385 - 550g)
gemfibrozil	0.10	Dix KJ, et.al. Drug Metab Dispos. 1999 Jan;27(1):138-46 Female Sprague-Dawley rats (10-12 weeks old).
glutarylsulphathiazole	42.00	Hirom PC, et.al. Biochem J. 1972 Oct;129(5):1071-7. Female Wistar albino rats (180-350g body wt.)
grepafloxacin	5.81	Sasabe H, et.al. J Pharmacol Exp Ther. 1998 Mar;284(3):1033-9. Male Sprague-Dawley rats weighing approximately 250 to 300g. Sasabe H,et.al. J Pharmacol Exp Ther. 1998 Feb;284(2):661-8. Male Sprague-Dawley (SD) weighing approximately 250 to 300g. Yamaguchi H, et. al. J Pharmacol Exp Ther. 2002 Mar;300(3):1063-9. Male Wistar rats, 200-240g. Yamaguchi H, et.al. Pharm Res. 2004 Feb;21(2):330-8. Male Wistar rats weighing 200-250g
hexafluorenium	34.00	Meijer DK, et.al. Eur J Pharmacol. 1971 May;14(3):280-5 Male Wistar rats weighing 200-250g
Hexahydrophthalylsulfathiazole	80.00	Hirom PC, et.al. Biochem J. 1972 Oct;129(5):1071-7. Female Wistar albino rats (180-350g body wt.)
Hippuric acid	0.00	Abou-el-makarem AA, Millburn P, et al Biochem. J.(1967)105, 1269
ID-6105	19.76	Yoo BI, et.al. Biol Pharm Bull. 2005 Apr;28(4):688-93 Male SD rats (230 - 250 g) . Yoo BI, et.a. Arch Pharm Res. 2005 Apr;28(4):476-82

Compounds	BE%	Reference
		Male SD rats (230 - 250 g).
Indocyanine Green	30.00	Jansen PL et.al. Am J Physiol. 1993 Sep;265(3 Pt 1):G445-52. Male Wistar rats, weighing 250-300g Kurisu H, et.al. Life Sci. 1991;49(14):1003-11. Sprague-Dawley Rat Verkade HJ, et.al. Gastroenterology. 1990 Nov;99(5):1485-92. Normal Wistar rats weighting 280-320 g. Sathirakul K, et.al. J Pharmacol Exp Ther. 1993 Jun;265(3):1301-12. Male SD rats weighing approximately 280 g. Takikawa H, et.al. J Gastroenterol Hepatol. 1998 Apr;13(4):427-32. Male Sprague-Dawley rats weighting approximately 270g. Hirom PC, et.al. Biochem J. 1972 Oct;129(5):1071-7. Female Wistar albino rats (180-350g body wt.) Tachizawa H, et.al. J Gastroenterol Hepatol. 2004 Sep;19(9):1016-22. Male Spraguc-Dawley rats 270g. Kimura T, et.al. Biol Pharm Bull. 1993 Nov;16(11):1140-5. Male Wistar rats weighing 200-300g Chan PK, et.al. J Toxicol Environ Health. 1981 Feb;7(2):169-79.
indomethacin	2.06	Kouzuki H, et.al. Pharm Res. 2000 Apr;17(4):432-8 SD rats of 302-368 g body weight.
iododoxorubicin (IODOX)	22.00	Edwards DM, et.al. Drug Metab Dispos. 1991 Sep-Oct;19(5):938-45. Male SD rats (mean weight 201 ± 6g). % of Dose (in total): 34; parent drug accounted for < 6% of that.
irinotecan (CPT-11) (lactone form)	7.34	Chu XY, et.al. J Pharmacol Exp Ther. 1997 Apr;281(1):304-14. Male SD rats weighing 250 to 300g. Arimori K, et.al. Pharm Res. 2003 Jun;20(6):910-7 Male Wistar rats from 280 to 340g. Itoh T, et.al. J Pharm Pharm Sci. 2004 Jan 23;7(1):13-8. Male Wistar rats, aged 6 to 7 weeks (180-230 g
J-104132	99.70	Kobayashi N, et.al. Pharm Res. 2003 Jan;20(1):89-95 Male SDRs (250-470 g).
lamotrigine	1.40	Maggs JL, et.al. Chem Res Toxicol. 2000 Nov;13(11):1075-81. Male Wistar rats (180-250g)
levofloxacin	9.04	Yamaguchi H, et. al. J Pharmacol Exp Ther. 2002 Mar;300(3):1063-9 Male Wistar rats, 200-240g. Yamaguchi H, et.al. Pharm Res. 2004 Feb;21(2):330-8. Male Wistar rats weighing 200-250g
Lissamine Fast Yellow	87.50	Bertagni P, et.al. J Pharm Pharmacol. 1972 Aug;24(8):620-4. Male and female Wistar albino rats (190-350 g). Hirom PC, et.al. Biochem J. 1972 Oct;129(5):1071-7. Female Wistar albino rats (180-350g body wt).
lithocholate (LC)	0.98	Takikawa H,et.al. Hepatology. 1991 Aug;14(2):352-60. Male SDRs weighing about 270g. % of dose (in total): 98% ± 1.6%. Parent drug accounted for 1% ± 1% of that.
lomefloxacin	4.26	Sasabe H, et.al. Biopharm Drug Dispos. 1999. Apr;20(3):151-8. Male SD rats weighing approximately 250-300g.
lopinavir	0.40	Kumar GN, et.al. Pharm Res. 2004 Sep;21(9):1622-30

Compounds	BE%	Reference
		Sprague-Dawley rats
loteprednol etabonate	4.84	Wu.W, F. Huang; J of pharmacy and pharmacology, 60(3),2008, 291-297
LTC4(leukotriene C4)	23.10	K.Niinuma,Y.Kato, et al American journal of physiology: Gastrointestinal & liver physiology, 1999 ;276(5 Pt 1)1153-1164 Denzlinger C, Grimberg M, Kapp A, Haberl C, WILMANNS W , British journal of pharmacology; 1991 102 (4),865-870, male Wistar rats(180-220 g)
LY110264	34.40	Wright W.E., Line V.D. Antimicrobial Agents & Chemotherapy(1980)17, 842-846. Male Wistar rats [HAP(WI)BR], weighing 350 to 500
LY112384	84.70	Wright W.E., Line V.D. Antimicrobial Agents & Chemotherapy(1980)17, 842-846. Male Wistar rats [HAP(WI)BR], weighing 350 to 500
LY126351	11.00	Wright W.E., Line V.D. Antimicrobial Agents & Chemotherapy(1980)17, 842-846. Male Wistar rats [HAP(WI)BR], weighing 350 to 500
LY78989	74.20	Wright W.E., Line V.D. Antimicrobial Agents & Chemotherapy(1980)17, 842-846. Male Wistar rats [HAP(WI)BR], weighing 350 to 500
LY85834	40.30	Wright W.E., Line V.D. Antimicrobial Agents & Chemotherapy(1980)17, 842-846. Male Wistar rats [HAP(WI)BR], weighing 350 to 500
LY87780	93.80	Wright W.E., Line V.D. Antimicrobial Agents & Chemotherapy(1980)17, 842-846. Male Wistar rats [HAP(WI)BR], weighing 350 to 500
LY88011	49.60	Wright W.E., Line V.D. Antimicrobial Agents & Chemotherapy(1980)17, 842-846. Male Wistar rats [HAP(WI)BR], weighing 350 to 500
LY89439	49.60	Wright W.E., Line V.D. Antimicrobial Agents & Chemotherapy(1980)17, 842-846. Male Wistar rats [HAP(WI)BR], weighing 350 to 500
merck compound A	30.00	Giuliano C, et.al. Xenobiotica. 2005 Oct-Nov;35(10-11):1035-54. Male Sprague-Dawley rats weighing 250-300g.
Meropenem	80.20	Yl.chan, MH.Chou, J Chromatogr A. 2002 Jun 28;961(1):119-24. Male specific pathogen-free Sprague–Dawley rats.
methadone	8.80	Baselt RC, et.al. Biochem Pharmacol. 1973 Dec 1;22(23):3117-20. Sprague-Dawley male rats (200 - 300 g).
methasquin (NSC 122870)	29.00	Rader JI, et.al. Cancer Res. 1971 Jul;31(7):964-9 CD males, 230 to 420 g
methotrexate	72.00	Masuda M, et.al. Cancer Res. 1997 Aug 15;57(16):3506-10. Male SDRs (250 - 300g). Lutz Fahrig, Helmut Brasch, et al, Cancer Chemother Pharmacol(1989)23, 156-160 Sasaki M, et.al. Mol Pharmacol. 2004 Sep;66(3):450-9 Male SD rats, 240-260g. CLtot: 12.7 ± 1.9 ml/min/kg; CLbiliary: 10.7± 1.7 ml/min/kg Chen C, et.al. Pharm Res. 2003 Jan;20(1):31-7. Male Sprague-Dawley rats, 15 weeks of age (385 - 550g) Griffin D, et.al. Cancer Chemother Pharmacol. 1987;19(1):40-1 Ueda K, et.al. J Pharmacol Exp Ther. 2001 Jun;297(3):1036-43. Male Sprague-Dawley rats

Compounds	BE%	Reference
		weighing 250 to 300 g. Bremnes RM, et.al. Cancer Res. 1989 May 1;49(9):2460-4 Male Wistar rats weighing 220-300 g. Steinberg SE, et.al. Cancer Res. 1982 Apr;42(4):1279-82. Female Sprague-Dawley rats weighing 175 to 250 g.
Methyl orange	55.00	O'reilly W.J., Pitt P.A. et al, Br. J. Pharmac (1971), 43, 167-179.
Methylphenyldipropylammonium iodide	17.00	Hughes R.D., Millburn P., et al, Biochem. J. (1973)136, 967-78.
Mitoxantrone	6.08	Yang XN, Morris ME. J OF PHARM SCI, vol 99 (5) Pages: 2502-2510, May 2010. Male Sprague–Dawley (SD) rats (300–430 g).
morphine	9.03	Roerig DL, et.al. Biochem Pharmacol. 1974 Apr 15;23(8):1331-9. Sprague-Dawley male rats (300 - 400g). % of dose (in total): 49.3 ± 3.6. Peterson RE, et.al. J Pharmacol Exp Ther. 1973 184(2):409-18. Male SD rats (325-450 g). % of dose (in total): 63. Parent drug accounted for 17.0 ± 2.3% of that. Smith DS, et.al. Biochem Pharmacol. 1973 Feb 15;22(4):485-92. Male SD rats (350-450 g). % of dose (in total): 64 ± 5. Parent drug accounted for 10% of that.
moxalactam (latamoxef)	20.50	Uchida K, et.al. J Pharmacobiodyn. 1985 Nov;8(11):981-8 Wistar strain male rats, 8 weeks of age. Mizojiri K, et.al. Antimicrob Agents Chemother. 1987 Aug;31(8):1169-76. Male Sprague-Dawley rats (weight, 250 to 320 g)
MX-68	84.00	Han YH, et.al. J Pharmacol Exp Ther. 1999 Oct;291(1):204-12. Male Sprague-Dawley rats (SDRs) weighing 250 to 300g.
N2-methyl-9-hydroxyolivacinium	2.20	Maftouh M, et.al. Xenobiotica. 1983 May;13(5):303-10 Male SD rats (300 - 350g).
nafenopin	4.00	Jedlitschky G, et.al. Biochem Pharmacol. 1994 Sep 15;48(6):1113-20. % of Dose (in total): 40; Parent drug accounted for 10% of that.
naftopidil	6.60	Niebch G, et.al. Arzneimittelforschung. 1991 Oct;41(10):1027-32. Male Sprague-Dawley rats (150-200g)
NAPAP	37.90	Hauptmann J, et.al. Biomed Biochim Acta. 1987;46(6):445-53.Wistar Rats of both sexes, body weight 260-340g. Hauptmann J, et.al. Pharmazie. 1991 Jan;46(1):57-8
napsagatran	61.00	Lavé T, et.al. J Pharm Pharmacol. 1999 Jan;51(1):85-91 Male rats (230 ± 290 g), SPF, RoRo albino
nelfinavir	0.05	Kageyama M, et.al. Biol Pharm Bull. 2005 Feb;28(2):316-22. Male Wistar rats (280-320 g). Amount excreted into bile: 0.359 ± 0.027 µg. Dose: 2.5 mg/kg.
nitrofurantoin	5.16	Wang X, et.al. Drug Metab Dispos. 2007 Feb;35(2):268-74. Female SD rats (220g)
N-Methylpyridinium iodide	0.80	Hughes R.D., Millburn P., et al, Biochem. J. (1973)136, 967-78
octreotide	50.00	Yamada T, et.al. Biol Pharm Bull. 1998 Aug;21(8):874-8

Compounds	BE%	Reference
		Male Sprague-Dawley rats weighing approximately 220g.; Yamada T, et.al. J Pharmacol Exp Ther. 1996 Dec;279(3):1357-64.; Male SDR (approximately 220g). CLtot: 10.53 ± 0.38 ml/minl/kg. CLbiliary: 4.15 ± 0.21 ml/min/kg.; Yamada T, et.al. Drug Metab Dispos. 1997 May;25(5):536-43. Male SDRs weighing 220g. CLtot: 12.63 ± 0.56 ml/minl/kg. CLbiliary: 7.44 ± 0.29 ml/min/kg. Lemaire M, et.al. Drug Metab Dispos. 1989 Nov-Dec;17(6):699-703.
Orthanilic acid	0.00	Abou-el-makarem M.M, Millburn P., et al Biochem. J.(1967)105, 1269
paclitaxel (taxol)	11.62	Monsarrat B, et.al. J Natl Cancer Inst Monogr. 1993;(15):39-46. Sprague-Dawley rats. Monsarrat B, et.al. Drug Metab Dispos. 1990 Nov-Dec;18(6):895-901. Luo G, Johnson S, et al, Drug Metab Dispos. J. (2010)38, 422-430
PAEB (procaine amid ethobromide) - not in other tables	32.20	Watkins JB 3rd, et.al. Drug Metab Dispos. 1987 Mar-Apr;15(2):177-83. Male Sprague-Dawley rats. Alterations in biliary excretory function by streptozotocin-induced diabetes
pancuronium	3.50	Upton RA, et.al. Anesth Analg. 1982 Apr;61(4):313-6 Male Sprague-Dawley rats, weighting 250-350g.
Paraquat di-iodide	0.50	Hughes R.D., Millburn P., et al, Biochem. J. (1973)136, 979-984
pefloxacin	3.94	Montay G, et.al. Antimicrob Agents Chemother. 1984 Apr;25(4):463-72. Male Wistar rats (200 to 300g)
penicillin G (benzylpenicillin)	20.78	Tsuji A, et.al. J Pharm Sci. 1983 Nov;72(11):1239-52. Male Wistar rats (240g). Ito K, et.al. Am J Physiol Gastrointest Liver Physiol. 2004 287(1):G42-9. Male SD rats weighing 240-300g. % of dose (in total): 31.7; Parent drug accounted for 50% of that.
penicillin V	29.50	Tsuji A, et.al. J Pharm Sci. 1983 Nov;72(11):1239-52. Male Wistar rats (240g).
Phenolphthalein	2.00	Millburn P,et al, Biochem. J. 1967; 105, 1275 Female Wistar albino rats (weighing 200 ± 10g.)
Phenolphthalein disulphate	74.00	Hirom PC, et.al. Biochem J. 1972 Oct;129(5):1071-7. Female Wistar albino rats (180-350g body wt.)
Phenolphthalein glucuronide	14.10	Itagaki S, et.al. Drug Metab Pharmacinet. 2003;18(4):238-44. Male SD rats (300 -350g). Amount excreted into bile in 1 hr: 311 ± 23.4 nmol/kg. Dose: 2.2 µmol/kg.
phenolsulfonephthalein (PSP, phenol red)	14.10	Itagaki S, et.al. Drug Metab Pharmacinet. 2003;18(4):238-44. Male SD rats (300 -350g). Amount excreted into bile in 1 hr: 311 ± 23.4 nmol/kg. Dose: 2.2 µmol/kg.
phenytoin (diphenylhydantoin)	0.40	Inaba T, et.al. Drug Metab Dispos. 1975 Mar-Apr;3(2):69-73.Wistar rats (250-330 g). % of Dose (in total): 28 or 54, Parent drug accounted for about 0.3 - 1.1% of that. El-Hawari AM, et.al. J Pharmacol Exp Ther. 1977 Apr;201(1):14-25. Male SD rats (180-280 g). % of Dose (in total) in 2 hr: 32, Parent drug accounted for 1.9 ± 0.2% of that.
PhIP	3.09	Dietrich CG, et.al. Carcinogenesis. 2001

Compounds	BE%	Reference
		May;22(5):805-11 Female wistar rats (200 - 250g).
pipecuronium	4.48	Bodrogi L, et.al. Arzneimittelforschung. 1980;30(2a):366-70.Female rats weighing 200 to 320g. % of Dose (in total): 6.36; Parent drug accounted for 69 - 72% of that.
pitavastatin	76.15	Hirano M, et.al. Mol Pharmacol. 2005 Sep;68(3):800-7 Male Sprague-Dawley rats weighing approximately 250 to 300g. Fujino H, et.al. Drug Metab Pharmacinet. 2002;17(5):449-56. Male Sprague-Dawley rats weighing approximately 250g
pravastatin	76.15	Akashi M, et.al. Hepatol Res. 2006 Feb 11,193-198 Male Sprague-Dawley rats weighting approximately 270g Fukumura S, et.al. Pharm Res. 1998 Jan;15(1):72-6 Male Sprague-Dawley rats (SDR) approximately 270g Marumo T, et.al. J Gastroenterol. 2004 Oct;39(10):981-7. Male Sprague-Dawley rats weighting approximately 270g Sasaki M, et.al. Mol Pharmacol. 2004 Sep;66(3):450-9 male Sprague-Dawley rats weighing approximately 240 to 260g. Takikawa H, et.al. J Gastroenterol Hepatol. 1998 Apr;13(4):427-32. Male Sprague-Dawley rats weighting approximately 270g. Ohashi M, et.al. Pharmacology. 2002 Sep;66(1):31-5. Ogasawara T, et.al. Hepatol Res. 2001 Jun;20(2):221-231 Male Sprague-Dawley rats weighing approximately 270g Niinuma K, Kato Y, et al, Am J Physiol. 1999 ;276(5 Pt 1)1153-1164. Fukuda H, Ohashi R, et al,Drug Metab Dispos. 2008 Jul;36(7):1275-82
probenecid	13.62	Conway W, et.al. J Pharm Sci. 1974 Oct;63(10):1551-4 Male SD rats weighting 420- 530g. Guarino AM, et.al. J Pharmacol Exp Ther. 1968 Dec;164(2):387-95. Male Sprague-Dawley rats, weighing 250 to 320g. % of Dose (in total): 85.5 ± 2.7, 57.9 ± 4.0, 25.4 ± 3.4. Parent drug accounted for 16.2%, 37.7% and 34.6% of that.
prostacyclin (PGI 2)	0.00	Taylor BM, et.al. J Pharmacol Exp Ther. 1980 Jul;214(1):24-30 Female SD rats (200 - 250g)
proxicromil	4.40	Smith DA, et.al. Eur J Drug Metab Pharmacokinet. 1983 8(3):225-32. CRCD rats. Amount excreted into bile: 110 µg. Dose: 10 mg/kg. Weight assumed to be 250 g.
PSC 833(Valspodar)	0.86	Song S, et.al. Drug Metab Dispos. 1998 Nov;26(11):1128-33. Female Sprague-Dawley rats (10 weeks of age, weighing 220-270g)
QMPB	0.00	Christensen A, et.al. Xenobiotica. 1990 Apr;20(4):417-34 female Sprague-Dawley rat, body wt 200g
ramatroban	16.00	Moriwaki T, et.al. Pharm Res. 2004 Jun;21(6):1055-64 SDR weighing 200-220g. % of dose (in total): 28.5 ± 2.6, Parent drug accounted for 56% of that.
R-benoxaprofen	0.70	Mohri K, et.al. Pharm Res. 2005 Jan;22(1):79-85

Compounds	BE%	Reference
		Male SD rats (250 - 300g)
R-carprofen	9.84	Kemmerer JM, et.al.J Pharm Sci. 1979 Oct;68(10):1274-80. Male rats (200-300g)
remikiren	34.60	Coassolo P, et.al. Xenobiotica. 1996 Mar;26(3):333-45 Male albino SPF rats (weight 280-320 g)
reproterol	4.09	Kucharczyk N, et.al. Arzneimittelforschung. 1981;31(12):2085-8. Male Charles River rats (165-275g). % of Dose (in total): 45.33 ± 4.62; Parent drug accounted for 1.7 to 13% of that.
R-grepafloxacin	4.43	Sasabe H, et.al. Biopharm Drug Dispos. 1999. Apr;20(3):151-8. Male SD rats weighing approximately 250-300g
Rhodamine 123	3.72	Kageyama M, et.al. Biol Pharm Bull. 2006 Apr;29(4):779-84.Male Wistar Rats, 300 ± 20g. Amount excreted into bile over 2 hr: 2.23 ± 0.06 µg. Dose: 0.2 mg/kg. Kageyama M, et.al. Biol Pharm Bull. 2005 Feb;28(2):316-22. Male Wistar rats: 280 -320g. Amount excreted into bile: 2.79 ± 0.37 µg. Dose: 0.2 mg/kg. Yumoto R, et.al. Drug Metab Dispos. 2001 Feb;29(2):145-51. Male Wistar rats weighing 230 to 300g. Kageyama M, et.al. Biol Pharm Bull. 2005 Jan;28(1):130-7. Male Wistar: 300 ± 20g. Amount excreted into bile over 2 hr: ~ 2000ng. Dose: 0.2 mg/kg.
ritonavir	3.40	Denissen JF, et.al. Drug Metab Dispos. 1997.Apr;25(4):489-501. SD rats (220-270g). % of Dose (in total) in 6 hr: 79.7; Parent drug accounted for 1.9% of that (Male). % of Dose (in total) in 6 hr: 41.6; Parent drug accounted for 12.7% of that (Female).
Rivaroxaban	48.40	Weinz C, Schwarz T, Kubitza D. et al. (2009). Drug Metab Dispos. 2009;37(5):1056-64.
rosuvastatin	56.90	Kitamura S, et.al. Drug Metab Dispos. 2008 Oct;36(10):2014-23. Male Sprague-Dawley rats (9 weeks old) . H.Fukuda, R.Ohashi, et al,Drug Metab Dispos. 2008 Jul;36(7):1275-82. Male Sprague-Dawley rats (Charles River Japan, Yokohama, Japan) weighing 200 to 250 g
Salicylic acid	4.40	H.Fukuda, R.Ohashi, et al,Drug Metab Dispos. 2008 Jul;36(7):1275-82. Male Sprague-Dawley rats (Charles River Japan, Yokohama, Japan) weighing 200 to 250 g
SB-265123	2.80	WARD K, et.al. Drug Metab Dispos. 1999 Nov;27(11):1232-41. male Sprague-Dawley rats weighing 290 to 350 g
S-benoxaprofen	3.00	Mohri K, et.al. Pharm Res. 2005 Jan;22(1):79-85 Male SD rats (250 - 300g)
S-carprofen	5.70	Kemmerer JM, et.al.J Pharm Sci. 1979 Oct;68(10):1274-80. Male rats (200-300g)
s-grepafloxacin	3.66	Sasabe H, et.al. Biopharm Drug Dispos. 1999 Apr;20(3):151-8. Male SD rats weighing approximately 250-300g
sitagliptin	16.39	Beconi MG, et.al. Drug Metab Dispos. 2007 Apr;35(4):525-32. Male SD rats (360 - 450g).
SK&F 110679	53.10	Davis CB, et.al. Drug Metab Dispos. 1994 Jan-Feb;22(1):90-8. Male Sprague-Dawley rats.

Compounds	BE%	Reference
SN-38 carboxylate	6.72	ltagaki S, Sasaki K, et al, J Pharm Pharm Sci. (2004)23;7(1), 8-13. Male Wistar rats, aged 6 to 7 weeks (180-230 g in weight).
SN-38 lactone	2.43	ltagaki S, Sasaki K, et al, J Pharm Pharm Sci. (2004)23;7(1), 8-13. Male Wistar rats, aged 6 to 7 weeks (180-230 g in weight).
SN-38-glucuronide carboxylate	7.00	ltagaki S, Sasaki K, et al, J Pharm Pharm Sci. (2004)23;7(1), 8-13. Male Wistar rats, aged 6 to 7 weeks (180-230 g in weight).
SN-38-glucuronide lactone	21.90	ltagaki S, Sasaki K, et al, J Pharm Pharm Sci. (2004)23;7(1), 8-13. Male Wistar rats, aged 6 to 7 weeks (180-230 g in weight).
Stilboestrol	2.92	Millburn P,et al, Biochem. J. 1967; 105, 1275 Female Wistar albino rats (weighing 200 ± 10g.) % of Dose (in total): 94; Parent drug accounted for 3% of that.
stilboestrol glucuronide	89.00	Millburn P,et al, Biochem. J. 1967; 105, 1275 Female Wistar albino rats (weighing 200 ± 10g.)% of Dose (in total): 100; Parent drug accounted for 89% of that.
Succinylsulphathiazole	33.00	Hirom P.C., Millburn P., et al, Biochem. J. (1972)129, 1071-1077, Female Wistar albino rats (180-350g body wt.)
sulfaethidole	18.50	Kekki M, et.al. J Pharmacokinet Biopharm. 1982 Feb;10(1):27-51. Male Sprague-Dawley rats weighting 356 ± 12 g
sulphanilic acid	0.69	McMahon KA, et.al. Food Cosmet Toxicol. 1969 Sep;7(5):497-500. Rats (250-350 g body weight M.M.Abou-el-makarem, P.millburn, et al Biochem. J.(1967)105, 1269
Tartrazine	19.11	Hirom P.C., Millburn P., et al, Biochem. J. (1972)129, 1071-1077, Female Wistar albino rats (180-350g body wt.) Gregson RH, et.al. J Pharm Pharmacol. 1972 Jan;24(1):20-4. Male and female Wistar albino rats, 190-210g Bertagni P, et.al. J Pharm Pharmacol. 1972. 24(8):620-4
taurocholate	96.00	Akashi M, et.al. Hepatol Res. 2006 Feb 11,193-198 Male Sprague-Dawley rats weighting approximately 270g Takikawa H, et.al. Hepatology. 1996 Mar;23(3):607-13. Male Sprague-Dawley rats (SDR) approximately 270g. Fukumura S, et.al. Pharm Res. 1998 Jan;15(1):72-6 Male Sprague-Dawley rats (SDR) approximately 270g. Kuipers F, et.al. J Clin Invest. 1988 May;81(5):1593-9 Wistar rats Jansen PL, et.al. Hepatology. 1987 Jan-Feb;7(1):71-6. Homozygous TM rats (200 to 250g) Bowmer CJ, et.al. Br J Pharmacol. 1984 Nov;83(3):773-82 Male Wistar albino rats (250-350g) Bode KA, et.al. Biochem Pharmacol. 2002 Jul 1;64(1):151-8 Male Wistar rats weighing about 180 - 220g. Meijer DK, et.al. Drug Metab Dispos. 1976 Jan-Feb;4(1):1-7. Male Wistar rats weighing about 275g. Watkins JB, et.al. Drug Metab Dispos. 1987 Mar-Apr;15(2):177-83. Male Sprague-Dawley rats.

Compounds	BE%	Reference
telithromycin	13.80	Yamaguchi S, et.al. Antimicrob Agents Chemother. 2006 Jan;50(1):80-7. Male SD rats, 270-280g. CLsys: 6.97 ± 0.22 L/hr/kg. CLbiliary: 4.41 ± 0.21 ml/min.
temazepam	0.50	Tse FL, et.al. J Pharm Sci. 1983 Mar;72(3):311-2 Male Wistar strain rats average weight 250g
temocaprilat	67.16	Takikawa H, et.al. Hepatol Res. 2002 Oct;24(2):136 Male Sprague-Dawley rats (270 g) . Ishizuka H, et.al. J Pharmacol Exp Ther. 1997 Mar;280(3):1304-11. Male Sprague-Dawley rats (7 weeks old). Ishizuka H, et.al. J Pharmacol Exp Ther. 1999 Sep;290(3):1324-30. Male Sprague-Dawley (SD) rats.
terbutaline	7.88	Eriksson H, et.al. Acta Physiol Scand. 1975 Sep;95(1):1-5 Mak SPF Sprague-Dawley mts, wetghing 250 -300g. CLtot: 5.2 ml/min/kg; CLbiliary: 0.41 ml/ming/kg
Tetraethylammonium bromide	0.50	Hughes R.D, Millburn P., et al, Biochem. J. (1973)136, 967-78
tetrahydrocannabinol	0.07	Widman M, et.al. Biochem Pharmacol. 1974 Apr 1;23(7):1163-72. Sprague-Dawley rats. % of Dose (in total) in 6 hr: 68, Parent drug accounted for 0.1% of that
thyroxine (T4)	3.46	Wong H, et.al. Toxicol Sci. 2005 Apr;84(2):232-42 Male Sprague-Dawley rats approximately 8-10 weeks old (~ 225-325g).
tolrestat	54.75	Cayen MN, et.al. Drug Metab Dispos. 1985 Jul-Aug;13(4):412-9. Male albino SD rats (200-250 g). % of Dose (in total): 73 in 4 hr; Parent drug accounted for 75% of that.
TPBE	0.80	Dow J, et.al. Xenobiotica. 1982 Oct;12(10):633-43 Male Sprague-Dawley rats of approx. 150g
TR-14035	29.40	Tsuda-Tsukimoto N, et.al. Pharm Res. 2006 Nov;23(11):2646-56. Male Sprague-Dawley rats weighing 250 to 320 g.
triamterene	5.50	Kau ST, et.al. Drug Metab Dispos. 1975 Sep-Oct;3(5):345-51. Male SD rats (200 - 250g
tributylmethylammonium (TBuMA)	33.30	Hong SS, et.al. Pharm Res. 2000 Jul;17(7):833-8. Male Sprague-Dawley rats, 7 to 8 weeks of age. Han YH, et.al. Drug Metab Dispos. 1999 Aug;27(8):872-9 Male Wistar rats (250-300g). Hong SS, et.al. Arch Pharm Res. 2005 Mar;28(3):330-4 Male Sprague-Dawley rats, 7 to 8 weeks of age. Neef C, et.al. Naunyn Schmiedebergs Arch Pharmacol. 1984 Dec;328(2):103-10.Male Wistar rats, weighing approximately 300g. Lee IK, et.al. Arch Pharm Res. 2002 Dec;25(6):969-72.Male Sprague-Dawiey rats (250-270g). Jansen PL, et.al. Hepatology. 1987 Jan-Feb;7(1):71-6. Wistar rats: 200-250g.
triethylmethylammonium(TEMA)	0.39	Hong SS, et.al. Pharm Res. 2000 Jul;17(7):833-8. Male Sprague-Dawley rats, 7 to 8 weeks of age. Neef C, et.al. Naunyn Schmiedebergs Arch Pharmacol. 1984 Dec;328(2):103-10.Male Wistar rats, weighing approximately 300g. Han YH, et.al. Drug Metab Dispos. 1999 Aug;27(8):872-9. Male Wistar rats (250-300g).
trifluoperazine	0.30	Schmalzing G, et.al. Xenobiotica. 1978 Jan;8(1):45-54

Compounds	BE%	Reference
		Male Wistar rats. 200-250g.
triiodothyroacetic acid	1.05	Rutgers M, et.al. Endocrinology. 1989 Jul;125(1):433-43. Male Wistar rats (approximately 200 g). % of Dose (in total): 42 ± 4; Parent drug accounted for less than 2.5% of that.
Trimethylphenylammonium iodide	0.70	Hughes RD, Millburn P. et al, Biochem. J. (1973)136, 967-78
trimetrexate	0.80	Wong BK, et.al. Drug Metab Dispos. 1990 Nov-Dec;18(6):980-6 Male SD rats (333 to 382g).
UK-224,671	28.90	Beaumont K, et.al. Eur J Pharm Sci. 2000 Nov;12(1):41-50 Male Sprague-Dawley rats
UK-240,455	23.20	Webster R, et.al. Xenobiotica. 2003 May;33(5):541-60 Male Sprague-Dawley rats (300-350g).
UK-427,857	40.00	Walker DK, et.al. Drug Metab Dispos. 2005 Apr;33(4):587-95 Male Sprague-Dawley rats (250g).
ulifloxacin (UFX)	9.10	Yagi Y, et.al. Drug Metab Pharmacokinet. 2003;18(6):381-9 Male SD rats aged 7 weeks.
valsartan	42.75	Yamashiro W, et.al. Drug Metab Dispos. 2006 Jul;34(7):1247-54 Male Sprague-Dawley (SD) rats (7-8 weeks old). H.Fukuda, R.Ohashi, et al,Drug Metab Dispos. 2008 Jul;36(7):1275-82 Male Sprague-Dawley rats (Charles River Japan, Yokohama, Japan) weighing 200 to 250 g
vecuronium	46.00	Upton RA, et.al. Anesth Analg. 1982 Apr;61(4):313-6 Male Sprague-Dawley rats, weighting 250-350g.
verlukast(MK-571)	17.75	Nicoll-Griffith DA, et.al. Drug Metab Dispos. 1995 Oct;23(10):1085-93 male SD rats (~ 350g)
Vinblastine	30.00	Kurihara H, Sano N and Takikawa H. 2005. 20:1069-1074.
vincristine (VCR)	42.60	Song S, et.al. Drug Metab Dispos. 1999 Jun;27(6):689-94 Female Sprague-Dawley (SD) rats weighing 220 to 270g. Castle MC, et.al. Cancer Res. 1976 Oct;36(10):3684-9. Male and female SD rats (200 to 250g).
Voreloxin	35.20	Evanchik MJ, et al. Drug Metab Dispos. 2009 Mar;37(3):594-601. male Sprague-Dawley rats, weighing 225 to 275 g
xamoterol	0.00	Mulder GJ, et.al. Xenobiotica. 1987 Jan;17(1):85-92 Male Wistar rats (body wt approx. 200g). % of Dose (in total): 40; No unchanged drug existed.
YM-13115	72.20	Matsui H, et.al. Antimicrob Agents Chemother. 1984 Aug;26(2):204-7 male SD rats (body weight, 200 to 250 g)

Appendix II. Binding data for P-gp inhibitors

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
LY335979	Digoxin	Caco-2	0.02	177	0.023	5	Choo et al, 2000
Elacridar	Prazosin	MDCK II-MDR1			0.05	1	Rautio et al 2006
LY335979	Abacavir	MDCK II-MDR1	0.07		0.05		Shaik et al 2007
Loperamide	quinidine	MDCK II-MDR1			0.1	3	Lumen et al 2010
Reserpine	Daunomycin	P388 lymphoma			0.14	0.002	Lan et al, 1996
Verapamil	vincristine	K562-MDR	0.2	1.7	0.179	0.2	Richter et al 2009
Elacridar	calcein	MDCK II-MDR1	0.3	10	0.273	1	Matsson P et al 2009
Elacridar	Irinotecan	MDCK II-MDR1	0.38	46	0.312	10	Luo et al, 2002
Elacridar	Digoxin	Caco-2			0.39	0.011	Tang et al 2002
Mefloquine	Daunomycin	P388 lymphoma			0.43	0.002	Lan et al, 1996
Dipyridamole	Daunomycin	P388 lymphoma			0.52	0.002	Lan et al, 1996
Itraconazole	calcein	MDCK-MDR1	0.6	3.1	0.581	0.1	Cook et al, 2009
Terfenadine	Daunomycin	P388 lymphoma			0.63	0.002	Lan et al, 1996
CP147478	Digoxin	Caco-2	0.14		0.75	5	Wandal et al, 1999
Reserpine	vinblastine	LLC-PK1/MDR1			0.97	2	Ekins et al, 2002
Cyclosporine	Prazosin	MDCK II-MDR1			0.98	1	Rautio et al 2006
Verapamil	Prazosin	MDCK II-MDR1			1.18	1	Rautio et al 2006
Gallopamil	vinblastine	Caco-2	1.63	4.1	1.308	1	Neuhoff et al, 2000
Nelfinavir	Digoxin	Caco-2	1.4	177	1.362	5	Choo et al, 2000

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
Tamoxifen	Daunomycin	P388 lymphoma			1.39	0.002	Lan et al, 1996
D-703	Digoxin	Caco-2	1.6	177	1.556	5	Pauli-Magnus et al, 2000
pumafentrine	calcein	K562-MDR	3.12	0.3	1.56	0.25	Richter et al 2009
CP99542	Digoxin	Caco-2	3.8		1.6	5	Wandal et al, 1999
Erlotinib	vincristine	K562-MDR	2	1.7	1.787	0.2	Richter et al 2009
Cyclosporin	Digoxin	Caco-2			0.46	0.011	Noguchi et al, 2009
CP114769	Digoxin	Caco-2	0.3		2	5	Wandal et al, 1999
Quinidine	Daunomycin	P388 lymphoma			2.05	0.002	Lan et al, 1996
Ketoconazole	Prazosin	MDCK II-MDR1			2.38	1	Rautio et al 2006
Chlorpromazine	Daunomycin	P388 lymphoma			2.41	0.002	Lan et al, 1996
Bromocriptine	calcein	LLC-PK1/MDR1			2.81		Ekins et al, 2002
Ketoconazole	Digoxin	Caco-2	1.2	177	1.167	5	Cook et al, 2009
CP117227	Digoxin	Caco-2	0.07		3	5	Wandal et al, 1999
Norverapamil	vinblastine	Caco-2	4.24	4.1	3.402	1	Neuhoff et al, 2000
Promethazine	Daunomycin	P388 lymphoma			3.45	0.002	Lan et al, 1996
Itraconazole	Digoxin	Caco-2	2	385	1.974	5	Cook et al, 2009
Carvedilol	Digoxin	Caco-2	4	385	3.949	5	Cook et al, 2009
Bromocriptine	vinblastine	LLC-PK1/MDR1			3.96	2	Ekins et al, 2002
Nicardipine	calcein	MDCK-MDR1	4.2	3.1	4.069	0.1	Cook et al, 2009
Spironolactone	Daunomycin	P388 lymphoma			4.14	0.002	Lan et al, 1996
Norgallopamil	vinblastine	Caco-2	5.46	4.1	4.381	1	Neuhoff et al, 2000

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
Mibefradil	Digoxin	Caco-2	1.2	177	1.167	5	Ekins et al, 2002
Progesterone	Daunomycin	P388 lymphoma			4.6	0.002	Lan et al, 1996
tolafentrine	calcein	K562-MDR	9.46	0.3	4.73	0.25	Richter et al 2009
Telmisartan	Digoxin	Caco-2	5	385	4.936	5	Cook et al, 2009
Amprenavir	quinidine	MDCK II-MDR1			5	3	Lumen et al 2010
Fluphenazine	Daunomycin	P388 lymphoma			5.52	0.002	Lan et al, 1996
Mibefradil	calcein	MDCK-MDR1	6	3.1	5.813	0.1	Cook et al, 2009
CP101556	Digoxin	Caco-2	0.6		5.9	5	Wandal et al, 1999
Ritonavir	Digoxin	Caco-2	3.8	177	3.696	5	Choo et al, 2000
Fentanyl	Digoxin	Caco-2	6.5	177	6.321	5	Ekins et al, 2002
ergocryptine	vinblastine	LLC-PK1/MDR1			6.43	2	Ekins et al, 2002
Amitriptyline	Daunomycin	P388 lymphoma			7.53	0.002	Lan et al, 1996
Saquinavir	Digoxin	Caco-2	6.5	177	6.321	5	Choo et al, 2000
Montelukast	Digoxin	Caco-2	8	385	7.897	5	Cook et al, 2009
Nicardipine	Digoxin	Caco-2	8	385	7.897	5	Cook et al, 2009
Verapamil	fexofenadine	Caco-2	8.44	150	7.913	10	Petri et al, 2004
Amiodarone	calcein	LLC-PK1/MDR1			5.78	1	Ekins et al, 2002
Tiapamil	vinblastine	Caco-2	12	4.1	9.645	1	Neuhoff et al, 2000
Ivermectin	Digoxin	Caco-2	10	177	9.725	5	Ekins et al, 2002
Lovastatin	Digoxin	Caco-2	10	177	9.725	5	Ekins et al, 2002
Mitomycin C	Digoxin	Caco-2	10	177	9.725	5	Ekins et al, 2002
Procainamide	Digoxin	Caco-2	10	177	9.725	5	Ekins et al, 2002
Carvedilol	vinblastine	Caco-2	13.7	4.1	11.017	1	Neuhoff et al, 2000

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
Desmethylazelastine	Daunomycin	LLC-PK1/MDR1	11.8	24	11.783	0.035	Katoh et al, 2000
ergocryptine	calcein	LLC-PK1/MDR1			12.2	1	Ekins et al, 2002
CP100356	Digoxin	Caco-2	0.11		13	5	Wandal et al, 1999
CP12379	Digoxin	Caco-2	0.7		13	5	Wandal et al, 1999
Desethylamiodarone	Digoxin	LLC-PK1/MDR1	25.2	11	25.143	0.025	Katoh et al, 2000
ergocristine	vinblastine	LLC-PK1/MDR1			13.33	2	Ekins et al, 2002
Nitrendipine	Digoxin	Caco-2	14	385	13.821	5	Cook et al, 2009
ergotamine	vinblastine	LLC-PK1/MDR1			14.25	2	Ekins et al, 2002
Gemcabene	calcein	MDCK-MDR1	15	3.1	14.531	0.1	Cook et al, 2009
Isradipine	Digoxin	Caco-2	15	385	14.808	5	Cook et al, 2009
Verapamil	calcein	MDCK-MDR1	30	3.1	29.063	0.1	Cook et al, 2009
Desethylamiodarone	Daunomycin	LLC-PK1/MDR1	15.4	24	15.378	0.035	Katoh et al, 2000
Felodipine	calcein	MDCK-MDR1	16	3.1	15.5	0.1	Cook et al, 2009
quinidine	Digoxin	MDCK II-MDR1			0.1	0.03	Lumen et al 2010
Ketoconazole	calcein	LLC-PK1/MDR1			24.9	1	Ekins et al, 2002
Azelastine	Daunomycin	LLC-PK1/MDR1	16	24	15.977	0.035	Katoh et al, 2000
PSC-833 (Valsopodar)	Digoxin	Caco-2	0.11		16		Wandal et al, 1999
Carvedilol	calcein	MDCK-MDR1	17	3.1	16.469	0.1	Cook et al, 2009
Repaglinide	Digoxin	Caco-2	17	385	16.782	5	Cook et al, 2009
Troglitazone	calcein	MDCK-MDR1	19	3.1	18.406	0.1	Cook et al, 2009

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
Amiodarone	Digoxin	LLC-PK1/MDR1	5.48	11	5.431	0.1	Katoh et al, 2000
Azithromycin	Digoxin	Caco-2	21.8	177	21.201	5	Ebrel et al, 2007
Conivaptan	calcein	MDCK-MDR1	22	3.1	21.313	0.1	Cook et al, 2009
vinblastine	Prazosin	MDCK II-MDR1			21.9	1	Rautio et al 2006
Amiodarone	Daunomycin	LLC-PK1/MDR1	22.5	24	22.467	0.035	Katoh et al, 2000
CP69042	Digoxin	Caco-2	2.3		23	5	Wandal et al, 1999
Loperamide	calcein	MDCK II-MDR1	26	10	23.636	1	Matsson P et al 2009
MK571	calcein	MDCK II-MDR1	26	10	23.636	1	Matsson P et al 2009
Miconazole	vinblastine	LLC-PK1/MDR1			26.36	2	Ekins et al, 2002
Felodipine	Digoxin	Caco-2	29	385	28.628	5	Cook et al, 2009
Diltiazem	calcein	MDCK-MDR1	30	3.1	29.063	0.1	Cook et al, 2009
clotrimazole	vinblastine	LLC-PK1/MDR1			29.92	2	Ekins et al, 2002
Isradipine	calcein	MDCK-MDR1	31	3.1	30.031	0.1	Cook et al, 2009
Troglitazone	Digoxin	Caco-2	31	385	30.603	5	Cook et al, 2009
Dipyridamole	Digoxin	LLC-PK1/MDR1	40	11	40		Kakumoto 2002
Ranolazine	calcein	MDCK-MDR1	34	3.1	32.938	0.1	Cook et al, 2009
Clarithromycin	Digoxin	Caco-2	4.1	177	3.987	5	Ebrel et al,2007
Ritonavir	calcein	MDCK-MDR1	36	3.1	34.875	0.1	Cook et al, 2009
Diltiazem	Digoxin	Caco-2	36	385	35.538	5	Cook et al, 2009
Midazolam	calcein	K562-MDR	73.9	0.3	36.95	0.25	Richter et al 2009
Erythromycin	vinblastine	LLC-PK1/MDR1			37.79	2	Ekins et al, 2002

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
Conivaptan	Digoxin	Caco-2	39	385	38.5	5	Cook et al, 2009
Thioridazine	calcein	MDCK II-MDR1	45	10	40.909	1	Matsson P et al 2009
Desmethylazelastine	Digoxin	LLC-PK1/MDR1	41.8	11	41.705	0.025	Katoh et al, 2000
ergocristine	calcein	LLC-PK1/MDR1			42.8	1	Ekins et al, 2002
Lansoprazole	calcein	K562-MDR	86.9	0.3	43.45	0.25	Richter et al 2009
clotrimazole	calcein	LLC-PK1/MDR1			44		Ekins et al, 2002
Saquinavir	calcein	MDCK-MDR1	46	3.1	44.563	0.1	Cook et al, 2009
Nifedipine	calcein	MDCK-MDR1	47	3.1	45.531	0.1	Cook et al, 2009
Omeprazole	calcein	MDCK-MDR1	54	3.1	52.313	0.1	Cook et al, 2009
Talinolol	calcein	MDCK-MDR1	48	3.1	46.5	0.1	Cook et al, 2009
Ranolazine	Digoxin	Caco-2	49	385	48.372	5	Cook et al, 2009
Indinavir	Prazosin	MDCK II-MDR1			50	1	Rautio et al 2006
Nifedipine	Digoxin	Caco-2	53	385	52.321	5	Cook et al, 2009
vinblastine	Digoxin	Caco-2			8.92	0.011	Tang et al 2002
Cortisol	Digoxin	Caco-2	55	177	53.489	5	Ekins et al, 2002
Tamoxifen	Digoxin	Caco-2	55	177	53.489	5	Ekins et al, 2002
Pantoprazole	calcein	K562-MDR	108	0.3	54	0.25	Richter et al 2009
Clarithromycin	calcein	MDCK-MDR1	57	3.1	55.219	0.1	Cook et al, 2009
Miconazole	calcein	LLC-PK1/MDR1			55.5	1	Ekins et al, 2002
Paroxetine	calcein	MDCK-MDR1	61	3.1	59.094	0.1	Cook et al, 2009
Pantoprazole	Digoxin	Caco-2	69	385	68.115	5	Cook et al, 2009
Omeprazole	vinblastine	Caco-2	89	4.1	71.411	1	Neuhoff et al,

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
							2000
fluvastatin	calcein	K562-MDR	151	0.3	75.5	0.25	Richter et al 2009
Desmethylcarvedilol	vinblastine	Caco-2	97.6	4.1	78.311	1	Neuhoff et al, 2000
Daunomycin	Digoxin	Caco-2	55	177	53.489	5	Ekins et al, 2002
Troleandomycin	vinblastine	LLC-PK1/MDR1			87.64	2	Ekins et al, 2002
imipramine	calcein	K562-MDR	180	0.3	90	0.25	Richter et al 2009
Alprenolol	calcein	K562-MDR	181	0.3	90.5	0.25	Richter et al 2009
digoxin	calcein	K562-MDR	189	0.3	94.5	0.25	Richter et al 2009
Captopril	calcein	MDCK-MDR1	100	3.1	96.875	0.1	Cook et al, 2009
Cimetidine	calcein	MDCK-MDR1	100	3.1	96.875	0.1	Cook et al, 2009
Losartan	calcein	MDCK-MDR1	100	3.1	96.875	0.1	Cook et al, 2009
Milameline	calcein	MDCK-MDR1	100	3.1	96.875	0.1	Cook et al, 2009
Chlorzoxazone	Digoxin	Caco-2	100	177	97.253	5	Ekins et al, 2002
colchicine	Digoxin	Caco-2	100	177	97.253	5	Ekins et al, 2002
Debrisoquine	Digoxin	Caco-2	100	177	97.253	5	Ekins et al, 2002
fexofenadine	Digoxin	Caco-2	100	177	97.253	5	Ekins et al, 2002
Paclitaxel	Digoxin	Caco-2	100	177	97.253	5	Ekins et al, 2002
S-Mephenytoin	Digoxin	Caco-2	100	177	97.253	5	Ekins et al, 2002
Tolbutamide	Digoxin	Caco-2	100	177	97.253	5	Ekins et al, 2002
ergotamine	calcein	LLC-PK1/MDR1			98.9	1	Ekins et al, 2002
ergometrine	vinblastine	LLC-PK1/MDR1			100	2	Ekins et al, 2002
4-hydroxycarvediolol	vinblastine	Caco-2	128	4.1	102.591	1	Neuhoff et al, 2000

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
ergocornine	calcein	LLC-PK1/MDR1			105.2	1	Ekins et al, 2002
desipramine	calcein	K562-MDR	221	0.3	110.5	0.25	Richter et al 2009
ergometrine	calcein	LLC-PK1/MDR1			115.5	1	Ekins et al, 2002
Chlorprothixene	calcein	MDCK II-MDR1	130	10	118.182	1	Matsson & par 2009
guanabenz	calcein	K562-MDR	250	0.3	125	0.25	Richter et al 2009
Losartan	Digoxin	Caco-2	144	385	142.154	5	Cook et al, 2009
Verapamil	Irinotecan	MDCK II-MDR1	234	46	191.838	10	Luo et al, 2002
Avasimibe	Digoxin	Caco-2	200	385	197.436	5	Cook et al, 2009
Talinolol	Digoxin	Caco-2	294	385	290.231	5	Cook et al, 2009
Sitagliptin	Digoxin	Caco-2	300	385	296.154	5	Cook et al, 2009
Sparfloxacin	Digoxin	Caco-2	300	385	296.154	5	Cook et al, 2009
dihydroergocryptine	calcein	LLC-PK1/MDR1			360.5	1	Ekins et al, 2002
Fluconazole	vinblastine	LLC-PK1/MDR1			400	2	Ekins et al, 2002
Levofloxacin	Digoxin	Caco-2	500	385	493.59	5	Cook et al, 2009
Meloxicam	Digoxin	Caco-2	500	385	493.59	5	Cook et al, 2009
Orlistat	Digoxin	Caco-2	500	385	493.59	5	Cook et al, 2009
dihydroergocristine	calcein	LLC-PK1/MDR1			511	1	Ekins et al, 2002
Etoposide	Digoxin	Caco-2			294	0.011	Tang et al 2002
Etoposide	Irinotecan	MDCK II-MDR1	1185	46	971.486	10	Luo et al, 2002
Dilevalol	vinblastine	Caco-2	1185	4.1	950.81	1	Neuhoff et al, 2000
Captopril	Digoxin	Caco-2	1000	385	987.17	5	Cook et al, 2009

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
					9		
Cimetidine	Digoxin	Caco-2	1000	385	987.179	5	Cook et al, 2009
Milameline	Digoxin	Caco-2	1000	385	987.179	5	Cook et al, 2009
Paroxetine	Digoxin	Caco-2	1000	385	987.179	5	Cook et al, 2009
dihydroergotamine	calcein	LLC-PK1/MDR1			1000	1	Ekins et al, 2002
Fluconazole	calcein	LLC-PK1/MDR1			1000	1	Ekins et al, 2002
Diacetolol	vinblastine	Caco-2	3520	4.1	2824.348	1	Neuhoff et al, 2000
Cyclosporin	Daunomycin	P388 lymphoma			0.038	0.002	Lan et al, 1996
quinidine	quinidine	MDCK II-MDR1			0.1	3	Lumen et al 2010
Norverapamil	Digoxin	Caco-2	0.3	177	0.292	5	Pauli-Magnus et al, 2000
Propafenone	Daunomycin	P388 lymphoma			0.44	0.002	Lan et al, 1996
Verapamil	Daunomycin	P388 lymphoma			0.69	0.002	Lan et al, 1996
Verapamil	Digoxin	Caco-2	1.1	177	1.07	5	Pauli-Magnus et al, 2000
Verapamil	vinblastine	Caco-2	1.48	4.1	1.188	1	Neuhoff et al, 2000
Reserpine	Digoxin	Caco-2			1.38	0.011	Tang et al 2002
Telithromycin	Digoxin	Caco-2	1.8	177	1.751	5	Ebrel et al, 2007
Loperamide	Digoxin	Caco-2	2.7	177	2.626	5	Ekins et al, 2002
Trifluoperazine	Daunomycin	P388 lymphoma			3.8	0.002	Lan et al, 1996
Sufentanil	Digoxin	Caco-2	4.2	177	4.085	5	Ekins et al, 2002
Cyclosporine	calcein	LLC-PK1/MDR1			4.66	1	Ekins et al, 2002

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
Diltiazem	Daunomycin	P388 lymphoma			5.41	0.002	Lan et al, 1996
Telmisartan	calcein	MDCK-MDR1	6	3.1	5.813	0.1	Cook et al, 2009
Fluoxetine	Digoxin	Caco-2	10	177	9.725	5	Ekins et al, 2002
Terfenadine	Digoxin	Caco-2	10	177	9.725	5	Ekins et al, 2002
Quinidine	calcein	MDCK-MDR1	11	3.1	10.656	0.1	Cook et al, 2009
Reserpine	calcein	LLC-PK1/MDR1			12.2	1	Ekins et al, 2002
Quinidine	Prazosin	MDCK II-MDR1			14	1	Rautio et al 2006
Roxithromycin	Digoxin	Caco-2	15.4	177	14.977	5	Eberl et al, 2007
dihydroergocristine	vinblastine	LLC-PK1/MDR1			16	2	Ekins et al, 2002
dihydroergocryptine	vinblastine	LLC-PK1/MDR1			19.82	2	Ekins et al, 2002
Erythromycin	Digoxin	Caco-2	22.7	177	22.076	5	Eberl et al, 2007
ergocornine	vinblastine	LLC-PK1/MDR1			24.5	2	Ekins et al, 2002
testosterone	calcein	K562-MDR	56.4	0.3	28.2	0.25	Richter et al 2009
Azelastine	Digoxin	LLC-PK1/MDR1	30	11	29.932	0.025	Katoh et al, 2000
Haloperidol	calcein	MDCK II-MDR1	39	10	35.455	1	Matsson P et al 2009
Nitrendipine	calcein	MDCK-MDR1	41	3.1	39.719	0.1	Cook et al, 2009
Indinavir	Digoxin	Caco-2	44	177	42.791	5	Choo et al, 2000
Midazolam	Digoxin	Caco-2	55	177	53.489	5	Ekins et al, 2002
Citalopram	Digoxin	Caco-2	58	385	57.256	5	Cook et al, 2009
vincristine	Digoxin	Caco-2			71.1	0.011	Tang et al 2002
Omeprazole	Digoxin	Caco-2	85	385	83.91	5	Cook et al, 2009
Avasimibe	calcein	MDCK-MDR1	100	3.1	96.875	0.1	Cook et al, 2009

Inhibitor	Substrate	Cell System	IC50 (μM)	Km (μM)	Ki (μM)	Subst Conc (μM)	Reference
caffeine	Digoxin	Caco-2	100	177	97.253	5	Ekins et al, 2002
Morphine	Digoxin	Caco-2	100	177	97.253	5	Ekins et al, 2002
Amprenavir	Prazosin	MDCK II-MDR1			100	1	Rautio et al 2006
Alfentanil	Digoxin	Caco-2	112	177	108.923	5	Ekins et al, 2002
dihydroergotamine	vinblastine	LLC-PK1/MDR1			119.4	2	Ekins et al, 2002
5-hydroxycarvediolol	vinblastine	Caco-2	188	4.1	151.087	1	Neuhoff et al, 2000
Erythromycin	calcein	LLC-PK1/MDR1	1000	0.5	333.333	1	Ekins et al, 2002
Troleandomycin	calcein	Caco-2			483.3	1	Ekins et al, 2002
Gemcabene	Digoxin	Caco-2	1000	385	987.179	5	Cook et al, 2009
Labetalol	vinblastine	Caco-2	2194	4.1	1760.403	1	Neuhoff et al, 2000

11. List of Conference Attended

1. PPI-NET Young Researchers Symposium. 7th April 2014. (Abstract, Poster). Imperial College London.
2. Exchange Fellowship Drug Discovery Workshop. 9-10th December 2013. Milton Keynes, UK.
3. Marie Curie Initial Training Network (ITN) conference "Environmental ChemOinformatics" (ECO). (Abstract, Poster). 15-18th September 2013. Hosted by Helmholtz-Zentrum München. Chiemsee, Germany.
4. UK-QSAR Meeting (**Awarded** best poster prize - short talk). 23th April 2013. Hosted by Unilever. UK.
5. AAPS student conference (Poster). University of Greenwich. 25th March 2013. London.
6. 4th RSC/SCI International Symposium on Ion Channels as Therapeutic Targets (Winner of student **bursary**). 18th -19th March 2012. Abington Hall, Cambridge, UK.
7. Protein-Protein Interactions International Conference: Emerging science and therapeutic potential. (Full-**bursary** by conference). Abstract and Poster. 16th-17th January 2013. Royal Society of Chemistry, London.
8. NSCCS meeting 2012. Poster (student **bursary**). 12th December 2012. Imperial College London.
9. MOE training course ("hands-on" applications training by CCG). 10th -11th December 2012. University of Manchester. UK.
10. Medway School of Pharmacy Postgraduate Poster Day (Poster). 5th December, University of Kent. UK.
11. MGMS Young Modellers' Forum. (Abstract, Poster, Short talk). 30th November 2012. London.
12. UK-QSAR (Autumn Meeting). Poster. 8th November 2012. Cambridge, UK. Hosted by Takeda.
13. Cutting edge approaches to drug design symposium (Abstract, Poster) – 26th April 2012. London.

14. UK-QSAR Meeting (Poster). 25th April 2012. Horsham, UK. Hosted by Novartis.

15. 15th International workshop on Quantitative Structure-Activity Relationships in environmental and health science (Abstract, Poster) 18-22 June 2012, Estonia.

16. Postgraduate Teaching and Learning Course and Delivery strategies. Educational Development Unit by University of Greenwich, UK. (15 hours), Jan-Feb 2012 – **Certificate**.

17. RSC/ChemSoc Meeting. 17th Novermber 2011. University of Greenwich. UK.

18. DMDG open meeting (DMDG full student **bursary**) 14-16th September 2011. University of Cambridge (Poster).

19. Forth SFB – Symposium (Grant). September 8-9th 2011 – University of Vienna, Austria. Paper published in the Symposium journal (Abstract, Poster).

20. Experiences of applying system biology. BBSRC Funded seminar. 19th October 2011 – Kings College London. Poster presentation.

21. Chemoinformatics practical training course (winner of full-**bursary**). 21–24th June 2011 University of Sheffield. UK.

22. UK-QSAR Meeting (Abstract, Poster). 26th May 2011. University of Manchester. UK.

23. Schrödinger interactive web-seminar series "protein modelling and docking methods" on 25th and 27th of October and 3rd ,8th ,10th ,15th of November 2011.

24. Postgraduate poster presentation day. University of Greenwich. (poster) presentation. 17th June 2011.

25. Maestro interactive web-seminar series "Data Analysis and custom tools for lead generation". 16th and 18th October 2010.

12. List of Publications

Estimation of Biliary Excretion of Foreign Compounds Using Properties of Molecular Structure. **2014**. Sharifi M., Ghafourian T. AAPS J. 16 (1) 65-78.

Karanjin interferes with ABCB1, ABCC1, and ABCG2. Michaelis M, Rothweiler F, Nerreter T, Sharifi M, Ghafourian T, Cinatl J. **2014**. Journal of Pharmacy & Pharmaceutical Sciences (US). 17(1) 92-105.

Testing of SNS-032 in a panel of human neuroblastoma cell lines with acquired resistance to a broad range of drugs. Dec **2013**. Löschmann N, Michaelis M, Sharifi M. *et al.* Translational Oncology. 6 (6) 685-696.

Prediction of P-glycoprotein inhibition constant using QSAR. Sharifi M., Ghafourian T. Eur Med Chem. Submitted on 05 March - **2014**.

Differential effects of the oncogenic BRAF Inhibitor PLX4032 (vemurafenib) andits progenitor PLX4720 on ABCB1 function. Michaelis M, Rothweiler F, NerreterT, van Rikxoort M, Sharifi M, Wiese M, Ghafourian T, Cinatl J. Journal of Pharmacy & Pharmaceutical Sciences (US). Accepted on 11 March - **2014**.

Physiologically based pharmacokinetic modeling to predict biliary excretion using OATPs. Sharifi M. Ghafourian T. **2014**. In progress.