JRC Scientific and Technical Reports

Review of QSAR Models and Software Tools for predicting Biokinetic Properties

Aleksandra Mostrag-Szlichtyng and Andrew Worth

EUR 24377 EN - 2010

JRC EUROPEAN COMMISSION

ihcp Institute for Health and Consumer Protection

The mission of the IHCP is to provide scientific support to the development and implementation of EU policies related to health and consumer protection.

The IHCP carries out research to improve the understanding of potential health risks posed by chemical, physical and biological agents from various sources to which consumers are exposed.

European Commission
Joint Research Centre
Institute for Health and Consumer Protection

Contact information
Address: Via E. Fermi 2749, 21027 Ispra (VA), Italy
E-mail: andrew.worth@ec.europa.eu
Tel.: +39 0332 789566
Fax: +39 0332 786717

http://ihcp.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC58570

EUR 24377 EN
ISBN 978-92-79-15854-4
ISSN 1018-5593
DOI 10.2788/94537

Luxembourg: Publications Office of the European Union

Printed in Italy

ABSTRACT

In the assessment of industrial chemicals, cosmetic ingredients, and active substances in pesticides and biocides, metabolites and degradates are rarely tested for their toxicologcal effects in mammals. In the interests of animal welfare and cost-effectiveness, alternatives to animal testing are needed in the evaluation of these types of chemicals. In this report we review the current status of various types of *in silico* estimation methods for Absorption, Distribution, Metabolism and Excretion (ADME) properties, which are often important in discriminating between the toxicological profiles of parent compounds and their metabolites/degradation products. The review was performed in a broad sense, with emphasis on QSARs and rule-based approaches and their applicability to estimation of oral bioavailability, human intestinal absorption, blood-brain barrier penetration, plasma protein binding, metabolism and. This revealed a vast and rapidly growing literature and a range of software tools.

While it is difficult to give firm conclusions on the applicability of such tools, it is clear that many have been developed with pharmaceutical applications in mind, and as such may not be applicable to other types of chemicals (this would require further research investigation). On the other hand, a range of predictive methodologies have been explored and found promising, so there is merit in pursuing their applicability in the assessment of other types of chemicals and products. Many of the software tools are not transparent in terms of their predictive algorithms or underlying datasets. However, the literature identifies a set of commonly used descriptors that have been found useful in ADME prediction, so further research and model development activities could be based on such studies.

LIST OF ABBREVIATIONS

ADME	Absorption, distribution, metabolism and excretion
ADMET	Absorption, distribution, metabolism, excretion and toxicity
ANN	Artificial neural networks
ASNN	Associative Neural Networks
BB	Blood/brain partitioning
BBB	Blond/brain barrier
BM	Bayesian modelling
BNN	Bayesian neural networks
BRANN	Bayesian regularized artificial neural network
CART	Classification and regression trees
CG	Conjugate gradient
CL_{tot}	Non-metabolic clearance (bile and urinary elimination)
CNS	Central nervous system
CoMFA	Comparative molecular field analysis
CoMSIA	Comparative molecular similarity index analysis
CV	Cross-validation
DB	Database
DIDB	Metabolism & Transport Drug Interaction Database (University of Washington)
F	Oral bioavailability
FA	Fractional absorption
FALS	Fuzzy adaptive least squares
FIRM	Formal inference-based recursive modelling
GA	Genetic algorithm
GI	Gastrointestinal
GRNN	General regression neural networks
GST	Glutathione S-transferase
HIA	Human intestinal absorption
HM	Homology modelling
HQSAR	Hologram quantitative structure-property relationship
HSA	Human serum albumin
ILP	Inductive logic programming
KL	Kohonen learning
kNN	k-Nearest neighbours
k-PLS	Kernel-partial least squares
LDA	Linear discriminant analysis
LIE	Linear interaction energy
LOO-CV	Leave-one-out cross validation
LR	Logistic regression
MD	Molecular docking
MLP-NN	Multilayer perceptron neural networks
MLR	Multiple linear regression
ML	Machine learning
MM	Molecular modelling
MO	Molecular orbital calculations
MT	Methyltransferase
NAT	N-acetyltransferase
NN	Neural networks
P	Pharmacophore

PBPK	Physiologically-based pharmacokinetic model/modelling
PLS	Partial least squares
PLS-DA	Partial least squares-discriminant analysis
PPB	Plasma protein binding
QM	Semi-empirical quantum-mechanical calculations
QMSA	Quantitative molecular similarity analysis
QSAR	Quantitative structure-activity relationship
QSMR	Quantitative structure-metabolism relationship
QSPkR	Quantitative structure-pharmacokinetic relationship
QSPR	Quantitative structure-property relationship
RBF-NN	Radial basis function neural networks
RF	Random forest technique
RMSE	Root mean squared error
RRM	Ridge regression modelling
SDEC	Standard deviation of error of calculations
SDEP	Standard deviation of error of predictions
SOM	Self-organising map
SOMFA	Self-organising molecular field analysis
SULT	Sulfotransferase
SVM	Support vector machine
SVR	Support vector regression
TPSA	Topological polar surface area
TVM	Trend vector model
UFS	Unsupervised forward selection
UGT	UDP-glucuronosyltransferase
VSA	Van der Waals surface area

CONTENTS

1. Introduction

The term ADME refers to **A**bsorption, **D**istribution, **M**etabolism and **E**xcretion, the four processes related to the toxicokinetic (pharmacokinetic) profile of the chemicals interacting with living organisms. Collectively, these processes determine the fate of the substance inside the body. The term ADMET is sometimes also used, especially in the pharmacological area, the express the overall profiling of ADME properties and **T**oxicological effects of a substance.

The development of methods for determining ADME properties, including in silico methods, is a large and rapidly evolving field. This report provides an introduction to the background biology, then reviews of the current status of available databases, software tools and literature models relevant to ADME prediction. The in silico methods cover a range of approaches, including, but not limited to, (Q)SAR models.

2. Biological background of ADME processes

Absorption is a complicated process governed by a wide variety of factors, including not only the intrinsic properties of the substance (molecular size, solubility ($logS_{aq}$), ionization constant (pKa) and octanol/water partition coefficient (logP) values), but also physiological conditions inside the organism (local pH, absorptive surface area), and activities of enzymes, transporters and carriers along the gastrointestinal (GI) tract. Absorption in the upper GI tract (in mouth and stomach) is minimal and occurs as a result of passive diffusion. Substances absorbed in mouth (despite enzymatic degradation processes) enter directly the systemic circulation; substances absorbed in the stomach (despite hydrolysis and biotransformation processes) go to the liver first and their actual bioavailability is usually limited by first-pass metabolism. The most intensive absorption takes place in the lower GI tract, especially via large mucous surface of small intestine. The predominant absorption mechanism there is passive diffusion, although large molecules may be taken up by pinocytosis. In the large intestine absorption is less efficient and occurs by passive diffusion or active transport (in case of electrolytes). The activity of gut microflora, enzymatic degradation processes and hepatic first-pass metabolism usually diminish the amount of parent molecule that enters systemic circulation.

Human intestinal absorption (HIA) is usually measured as the percentage of the dose that reaches the portal vein after passing the intestinal wall (%HIA) and is a basis of most *in silico* absorption models. The percentage of the dose that remains after absorption and first-pass hepatic metabolism is defined as the oral bioavailability (F) of the compound. In other words, bioavailability describes the passage of a substance from the site of absorption into the systemic circulation and is usually not equivalent to the amount of a substance absorbed.

Once a compound enters the systemic circulation, it is distributed inside the body. This **distribution** process is governed by two main factors, namely the permeability of a substance between blood and particular tissues and the affinity of a substance to bind with tissues and plasma proteins.

One of the most important tissue/blood partitioning coefficients is blood/brain (BB) partition coefficient, usually expressed as logBB and defined as the ratio of substance concentration in blood to its concentration in brain. The passage of compounds across the blood/brain barrier (BBB), an important determinant of neurotoxicity, is based mainly on passive diffusion across the BBB membrane. However active transport also may be important. For nutrients and

endogenous compounds, such as amino acids, monocarboxylic acids, amines, hexoses, thyroid hormones, purine bases and nucleosides, several transport systems regulating the entry of the respective compound classes into the brain have been identified. In addition, there is evidence that active efflux pumps like the multidrug transporter P-glycoprotein (P-gp) on the luminal membrane of the brain capillary endothelial cells serve to impede the entry of hydrophobic compounds into the brain.

Compounds in the blood may exist in bound or unbound form. The protein binding of a substance influences the half-life inside the body and the bound fraction often serves as a reservoir from which the substance is slowly released to the unbound form. Unbound substances cross membrane barriers more readily, and may be metabolised and/or excreted. Hence the percentage of plasma protein binding (%PPB) is one of the key determinants in distribution. The most abundant protein in blood plasma is human serum albumin (HSA) accounting for about 60% of the total plasma protein. Since HSA is capable of binding diverse molecules, it significantly affects the overall %PPB.

Metabolism (**biotransformation**) is one of the main factors influencing the fate and toxicity of a chemical. Metabolism includes a set of chemical reactions (so-called metabolic pathways) inside the organism, which generally convert xenobiotics into more polar and more easily excreted (i.e. less toxic) forms. However, in some cases metabolism may lead to the formation of toxic metabolites or/and intermediates. Traditionally biotransformation is divided into two main phases - phase I and phase II. Phase I, the so-called functionalisation phase, has a major impact on lipophilic molecules, rendering them more polar and more readily excretable. In phase II, often referred to as detoxification, such functionalised moieties are subsequently conjugated with highly polar molecules before they are excreted. Both phases are catalysed by specific enzymes which are either membrane-bound (microsomal proteins) or present in the cytosol (cytosolic or soluble enzymes). The superfamily of cytochrome P450 (CYP450; also termed heme-thiolate protein P450) enzymes, including more than 70 families of proteins, catalyses the oxidative (and sometimes reductive) phase I metabolic reactions of diverse compounds. Phase II metabolism is governed by various enzymes acting on different types of molecules. The most significant among them are glutathione S-transferase (GST), methyltransferase (MT), N-acetyltransferase (NAT), sulfotransferase (SULT) and UDP-glucuronosyltransferase (UGT). Besides phase I and phase II metabolism, the liver causes specific pre-systemic (first-pass) effects, especially following the oral intake. In addition, phase III metabolism refers to the excretion of metabolites from cells with efflux transporters.

Excretion is the process of eliminating waste metabolic products, the major route of which is renal (urinary) excretion via the kidneys. The major non-metabolic routes of clearance (CL_{tot}) include bile and urinary elimination of unchanged compounds. The excretion with sweat, faeces and expired air as well as the ability of compounds to be excreted into breast milk and transferred to neonates may also be significant.

3. Literature reviews on the modelling of ADME properties

A detailed investigation of ADME properties is important during the drug discovery process, where it is used to optimize the bioavailability of drug candidates. However, toxicokinetic studies are also useful in toxicological investigations. Despite several difficulties in the modelling of ADME (e.g. low availability and/or quality of experimental data, complexity of physiological mechanisms inside the organisms), a large number of computational techniques have been developed. It is impossible to present here a comprehensive review of the literature on the modelling of ADME and ADME-related properties. As an illustration of the vastness

of the ADME literature, Table 1 lists the major reviews and expert opinions on *in silico* ADME prediction tools that have been published only during the last five years (2005-2010).

4. Databases and literature datasets

Although a wide range of diverse molecules have been screened in terms of their ADME properties, mainly to satisfy the needs of the pharmaceutical industry, relatively few data are publicly available. The majority of information on drug candidates are proprietary. Furthermore, ADME data for other types of chemicals (e.g. food additives, environmental pollutants, industrial chemicals, pesticides, etc.) are scarce. Thus, for the purpose of developing new ADME models, limited information is available. It is also unclear whether models developed for pharmaceuticals are applicable to a broader range of compounds, since pharmaceuticals are designed to be bioavailable and bioactive.

A list of available databases suitable for the development of QSARs for ADME properties is given in Table 2. One of them is **WOMBAT-PK 2009**, the clinical pharmacokinetics database of top selling drugs, provided by Sunset Molecular (http://www.sunsetmolecular.com/). It includes information about over 13 000 clinical pharmacokinetic measurements for 1230 molecules (1230 unique SMILES) and is being constantly expanded (over 100 drugs are planned to be added in 2010). All WOMBAT-PK 2009 drugs are represented (if possible) in neutral species. The searchable categories of WOMBAT-PK 2009 database include, among others, percentage oral bioavailability (for 818 drugs), percentage plasma protein binding (for 1006 drugs), percentage urinary excretion (for 811), qualitative blood brain barrier permeability (for 519 drugs) and phase I metabolizing enzymes (for 511 drugs). The **Metabolism & Transport Drug Interaction Database** (DIDB) has been developed by the University of Washington scientists (http://www.druginteractioninfo.org/). It contains *in vitro* and *in vivo* information on drug interactions in humans and provides pharmacokinetic profiles of drugs. The **Metabolite^{TM}** Database provided by Symyx (http://www.symyx.com/) indexes paths and schemes of biotransformation for xenobiotics and medicinal drugs and collects experimental data from *in vivo* and *in vitro* studies. **ADME DB**, a database provided by Fujitsu (http://www.fqs.pl/), contains data on interactions of substances with drug metabolizing enzymes and drug transporters. It includes information on ADME properties (e.g. CYP and other phase I and phase II enzymes) as well as interactions between drugs.

Among freely available databases, two are of importance (Table 2). The **ADME-AP** database developed by Bio Info & Drug Design (http://xin.cz3.nus.edu.sg/group/admeap/admeap.asp/) (Sun et al., 2002), provides data on diverse ADME-associated proteins including physiological function of each protein, pharmacokinetic effects, ADME classification, direction and driving force of disposition, location and tissue distribution, substrates, synonyms, gene name and protein availability in other species. The **PK/DB database** (http://www.pkdb.ifsc.usp.br/) includes 1203 compounds with respect to 2973 pharmacokinetic measurements (Moda et al., 2008). This database also includes five models for *in silico* ADME prediction (human intestinal absorption, human oral bioavailability, plasma protein binding, blood/brain barrier permeability and water solubility).

Numerous datasets published recently in the literature are also of importance as far as the modeling of ADME properties is concerned. They can be used for a wide range of predictive purposes, e.g. for human intestinal absorption, human oral bioavailability, plasma protein binding, blood brain barrier permeation and metabolic pathways modeling (Table 3).

5. Software

Tables 4 and 5 indicate the extensive range of software tools for the purpose of ADME and ADME-related predictions. The vast majority of available software tools are commercial. The tools differ greatly in terms of their capabilities and applications.

Some software, e.g. ACD/PhysChem Suite, ASTER, EPISUITE, ClogP (Table 4) were designed to perform the predictions of basic physicochemical properties (e.g. ionization constant pKa, octanol/water partition coefficient logP, distribution coefficient logD or aqueous solubility $\log S_{aq}$). The best accuracy attained in physicochemical property prediction is close to that of measured data. The only approach that promises to improve the predictive accuracy of such models seems to be consensus modelling, in which the results of multiple models are combined.

The importance of physicochemical property prediction is that the estimated data often serve as inputs to models of key ADME properties, such as gastrointestinal absorption, BBB permeability, oral bioavailability and plasma protein binding. Software tools such as Know-it-All, ADME Boxes, and ADMET Predictor (Table 5) generate physicochemical property predictions and use them in further ADME modeling.

In addition to structure-based models, there is a trend towards developing more sophisticated, mathematical PBPK models (Table 5). In these tools, *in vitro* and/or *in vivo* ADME data are integrated with the results of QSAR/QSPR models (e.g. for percentage plasma protein binding or blood/brain barrier penetration) for organism-based ADME modelling. Examples of such software tools include GastroPlus and Cloe which mimic the processes inside living organisms.

Simcyp (http://www.simcyp.com/) is a proprietary PBPK simulator that provides a platform for modelling the ADME properties of drugs and their metabolites, as well as drug-drug interactions, in virtual patient populations (Jamei et al, 2009). By predicting inter-individual variability, it can be used to identify people at the extreme risks arising from both oral and non-parenteral routes (lungs and skin) of drug administration/exposure. The populations included are: Healthy Volunteers, North European Caucasians, Japanese, Cirrhotic (different degree), Renal Impairment (different degrees), Obese (different levels), and all paediatric age groups. A Bayesian based parameter estimation module can be used to predict individual as well as population parameters. Various QSAR-based predictors are included to predict ADME parameters if measured data are not available. Simcyp is based on and includes a database of demographic, physiological, genomic and *in vitro* biochemical data. It has been developed by a consortium of pharmaceutical companies, academic institutes and regulatory authorities. In addition, as a module to the Simcyp Population-based ADME Simulator, Simcyp Rat is a 'virtual animal' for predicting drug kinetics in rats. Simcyp is a unique and comprehensive tool, and although is has been developed to support the safety assessment of drugs and their metabolites, it would be worth investigating for its applicability in dietary risk assessment.

6. Types of *in silico* modelling approaches identified in the literature

Literally thousands of ADME models have been published in the scientific literature during the last ten years. These models can be divided into a few categories of modelling approaches. The selection of the most useful approach depends on the aims of investigation and is usually driven by the availability of necessary input data, as well as by the level of information

needed as an output (e.g. high-throughput screening of numerous compounds or detailed analysis of particular metabolic reaction).

The simplest approach is based on **rules-of-thumb and structural alerts**. Their main advantages, i.e. simplicity and transparent interpretability, make them very useful for fast screening of large datasets. As far as ADME-related endpoints are concerned, several rules-of-thumb have been developed (Table 6), especially for assessing the likelihood of human intestinal absorption, blood/brain barrier penetration and plasma protein binding. Structural alerts have been identified mainly for metabolism-related issues, but also for human intestinal absorption (Raevsky et al., 2002). Models in this category are suitable for routine assessments by non-specialists, especially when rough approximations are sufficient.

Another approach to ADME modelling is **data-based modelling**. This includes conventional QSAR/QSPR and the application of different statistical algorithms, from relatively simple linear multivariate methods, such as Multiple Linear Regression (MLR), Partial Least Squares (PLS) and Linear Discriminant Analysis (LDA) to sophisticated nonlinear ones, such as Artificial Neural Networks (ANN). They are usually combined with learning methods such as Genetic Algorithms (GAs), Support Vector Machines (SVMs), Inductive Logic Programming (ILP), Bayesian Modelling (BM) and Self-Organizing Maps (SOMs). In addition, the approach of Hologram Quantitative Structure-Property Relationship (HQSAR) has been applied to ADME modelling. This technique is based on the arrangement of molecular fragments in a molecular hologram which allows three-dimensional information to be obtained from two-dimensional input structures (Wang et al., 2006; Moda et al., 2007a). Models in this category may be suitable for routine assessments by non-specialists, provided that a software implementation of the model is available.

QSARs for ADME properties tend to be local models, i.e. are based on small, homogenous data sets, with reliable predictions being obtained for the compounds falling within the model's applicability domain. Relatively few models have been developed on structurally diverse datasets containing more than 100 compounds. However, the accuracy of predictions across structurally diverse datasets can be improved by the application of **consensus modelling**, which transfers the strengths of multiple single models to a final consensus one. This approach has been demonstrated, for example in the modelling of blood/brain barrier penetration (Zhang et al., 2008a) and total clearance (Yap et al., 2006).

To obtain detailed information on the mechanisms of interaction between molecules, **similarity-based molecular modelling** may be useful. The methods within this category are used mainly in metabolism-related studies, especially for assessing the role of cytochrome P450 or identifying reaction sites (atoms) on particular enzyme substrates. Such methods include **3D-QSAR (e.g. Comparative Molecular Field Analysis, CoMFA); quantitative molecular similarity analysis (QMSA)**, based on experimental data or computed molecular descriptors; **pharmacophore modelling** and **docking**. Models in this category tend to require highly specialised modelling expertise, and as such are not suitable for routine assessments by non-specialists.

This review of literature-based ADME models given focuses on the conventional QSAR/QSPR approach (data-based modelling category), which could be useful for dietary risk assessment purposes. Given the large number of QSAR studies published and the wide variety of ADME properties (see Tables 7-12 for summary), the description is limited to a few illustrative examples, focusing on key ADME properties: human intestinal absorption (predicted as percentage fractional absorption, [%FA] or percentage human intestinal absorption [%HSA]), oral bioavailability (classification models), blood/brain barrier permeability (logBB and classification models), plasma protein binding (human serum

albumin [HSA] binding or percentage plasma protein binding [%PPB]) and excretion (total renal clearance [CL_{tot}]). Since conventional QSARs for metabolism prediction are highly limited in terms of their applicability, Table 11 also includes other types of literature biotransformation models (e.g. 3D-QSAR, pharmacophore modelling, docking).

6.1. Literature models for human intestinal absorption

Literature models for human intestinal absorption are summarised in Table 7.

Recently, an extensive dataset (Table 3) for human intestinal absorption was reported by Hou et al. (2007c). The authors proposed a classification Support Vector Machine (SVM) model of categorising compounds into high (%FA > 30%) or low fractional absorption (%FA < 30%) classes. The input data set of 578 structural diverse drug-like molecules, was split into training and test sets, including 480 and 98 molecules, respectively. Ten SVM classification models were generated by the LIBSVM software developed by Chang and Lin (freely available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/). The best model obtained gave satisfactory predictions for the training set (97.8% for the poor-absorption class and 94.5% for the good-absorption class). When the model was validated against an external test set, 100% of the poor-absorption class and 97.8% of the good-absorption class compounds were correctly classified. In the construction of classification models the authors considered 10 molecular descriptors, namely: the topological polar surface area (TPSA), the octanol/water partition coefficient (logP), the apparent partition coefficient at pH 6.5 (logD6.5), the number of violations of the four Rule-of-Five rules developed by Lipinski (Nrule-of-5), the number of hydrogen bond donors (nHBD), the number of hydrogen bond acceptors (nHBA), the intrinsic solubility (logS), the number of rotatable bonds (nrot), the molar volume (MV), and the molecular weight (MW). TPSA was calculated using the parameters originally proposed by Ertl et al. (2000). LogD was estimated based on the predicted logP and pKa calculated by ACDLABS 9.0. The remaining molecular descriptors were calculated using ACDLABS 9.0 (http://www.acdlabs.com/). The comparison of the models which were based on individual molecular descriptors showed that two of them (TPSA and logD6.5) are the most significant for the prediction of intestinal absorption. However, the best SVM model included seven descriptors, namely nHBD, logD6.5, MW, MV, TPSA, Nrule-of-5, and N+. The authors concluded that the size of the training set and its unbalanced nature have a great importance for predictivity of SVM classification models. A large data set is necessary for the model's stability and the ratio of poor-absorption class and good-absorption class compounds should be balanced to avoid model bias. Although the study of Hou et al. (2007c) was based on drug-like molecules, it indicates a modelling approach that might be useful for further research aimed at developing classification models of HIA for large sets of other types of compounds. Moreover, all data (%FA values and molecular descriptors) on the 578 studied compounds are available in the supporting materials provided with the paper and can be utilized by other scientists.

Another classification-based approach for human intestinal absorption modelling was proposed by Sun (2004). The authors used atom types as generic molecular descriptors, which allowed them to avoid making prior assumptions about which properties possibly related to the predicted endpoint. In total, 218 atom types were identified, including 88 types of C, 7 types of H, 55 types of N, 31 types of O, 8 types of halides, 23 types of S, and 6 types of P. In order to establish a qualitative model for HIA, the authors applied PLS-Discriminant Analysis (PLS-DA) to 169 drug molecules originally collected by Zhao et al. (2001) and classified them as "good" (absorption > 80%) “medium” (80% • absorption > 20%) or “poor” (absorption • 20%), according to the percentage human intestinal absorption (%HIA). A five-component PLS-DA model separated very well all 169 compounds. The goodness-of-fit was

expressed by regression coefficient (R^2) value of 0.92, and the model's predictivity by cross-validated regression coefficient (Q^2) of 0.79. In case only poorly absorbed compounds need to be identified, a three-component PLS-DA model would be sufficient, as it separated the compounds of less than 20% absorption with R^2 of 0.94 and Q^2 of 0.86. Although very often it is desirable to precisely predict, in many cases is sufficient, especially when fast recognition of the well/poorly absorbed molecules in large and diverse datasets is needed. The approach of Sun (2004) could be particularly useful in cases when it is sufficient to obtain a qualitative classification, rather than a precise estimate of the %HIA.

Using the Multicase approach, Klopman et al. (2002) compiled one of the largest datasets of 467 drug molecules for human intestinal absorption using various sources. The data were split into training and external prediction sets consisting of 417 and 50 molecules, respectively. In order to determine the molecular descriptors, both promoting and preventing HIA, the authors utilized the CASE program (http://www.multicase.com/). The occurrence of each structural fragment identified by CASE was subsequently used in a multiparameter linear equation of human intestinal absorption ($HIA = c_0 + \bullet c_i G_i$, where c_0 is a constant, c_i are the correlation coefficients and G_i is the presence (1) or absence (0) of a certain structural fragment). The final QSAR model was based on 37 descriptors: 36 statistically significant structural descriptors identified by CASE analysis and one important physicochemical parameter – the number of hydrogen bond donors (H_{donors}). The QSAR model was validated both internally (by multiple cross-validations) and externally (on the independent set of 50 drugs not included in the procedure of building the model). The final model displayed good statistics (correlation coefficient R^2 = 0.79, standard deviation error = 12.34%) and good predictive power (cross-validated R^2 for the external test set = 0.79). This study is useful in that it points to explicit substructures with negative (e.g. quaternary nitrogens, SO_2 groups connected to an aromatic ring) and positive impact for HIA, although is should be noted that that the training set was biased towards high absorption values. For the practical application of this model, it would first be necessary to rebuild it by using the using the Multicase software. The dataset would also need to be requested from Multicase (Klopman, 2002).

Abraham and collaborators (Zhao et al., 2001) proposed Linear Free Energy Relationship approach, based on the Abraham General Solvation Equation, to model human intestinal absorption. The authors created a %HIA dataset including 169 compounds. The model equation (%HIA = 92 + 2.94E + 4.10S + 10.6V – 21.7A – 21.1B), derived by stepwise MLR, was based on a set of five molecular descriptors, called Abraham descriptors, namely: excess molar refraction (E), solute polarity/polarizability (S), the McGowan characteristic volume (V), solute overall acidity (A) and basicity (B). The Abraham descriptors can be calculated by AbSolv program (previously available via ADME Boxes, Pharma Algorithms; now *via* ACD/ADME Suite, ACD Labs, http://www.acdlabs.com/). The model yielded the following statistics: correlation coefficient R^2 = 0.74, standard deviation error s = 14. The model indicated that the most significant descriptors in HIA prediction are hydrogen bond acidity and basicity as well as the McGowan volume – increasing the volume and decreasing the polarity of the molecule should enhance HIA. The ionisation state of acids or bases had no statistically significant impact on %HIA. The Abraham and Zhao approach was useful in identification of significant properties influencing the absorption processes and could be applied in further research aimed at the development of QSARs for quantitative prediction of HIA.

Another approach to human intestinal absorption predictions is based on the VolSurf methodology developed by Cruciani and co-workers (2000a,b). The VolSurf procedure allows to automatically convert the relevant information present in 3D molecular fields into few quantitative numerical molecular descriptors, which are more easily used and interpreted. The

VolSurf methodology is simple to apply as it is fast and fully automated. The descriptors are calculated using the GRID program (Molecular Discovery, http://www.moldiscovery.com/soft_grid.php/). In a study by Cruciani et al. (2000b), the VolSurf descriptors were used to model human absorption prediction. The authors defined a descriptor called "integy moment", analogous to the dipole moment, which discriminates between polar and non-polar parts of a molecule. High integy moments as well as hydrophobicity were found to be positively correlated with HIA. Conversely, a detrimental effect on absorption was correlated with polarity and a high concentration of polar interaction sites on the molecular surface.

The Jurs Research Group (Pennsylvania State University) has developed a wide range of diverse molecular descriptors which can be applied in predicting human intestinal absorption. For the purpose of such predictions, Wessel et al. (1998) identified the six most significant variables, namely: the cube root of gravitational index (connected with the size of molecule), the normalised 2D projection of the molecule on the YZ plane (SHDW-6, connected with the shape), the number of single bonds (NSB, connected with the flexibility), the charge on donatable hydrogen atoms (CHDH-1, connected with the hydrogen-bonding properties), the surface area multiplied by the charge of hydrogen bond acceptor atoms (SCAA-s, connected with the hydrogen-bonding properties) and the surface area of hydrogen bond acceptor atoms (SAAA-2, connected with the hydrogen-bonding properties). The authors used a set of 86 drugs with measured values of %HIA and applied a Genetic Algorithm with a Neural Network (GA-NN) technique to develop the model. The calculated %HIA model achieved good statistics with Root Mean Square Errors (RMSE) of 9.4%HIA units for the training set, 19.7%HIA units for the cross-validation (CV) set, and 16.0%HIA units for the external prediction set. All descriptors identified by Wessel et al. (amongst others) have been implemented in the **ADAPT** (Automated Data Analysis and Pattern Recognition Toolkit) software (available at http://research.chem.psu.edu/pcjgroup/adapt.html/). This is thus a useful research tool for the development of other QSAR models for human intestinal absorption.

6.2. Literature models for human oral bioavailability

Literature models for human oral bioavailability are summarised in Table 8.

A Support Vector Machine (SVM) approach for modelling human oral bioavailability has been proposed by Ma et al. (2008). The model was trained on 690 molecules by using a Support Vector Machine (SVM) method combined with Genetic Algorithm (GA) for feature selection and Conjugate Gradient (CG) for parameter optimization (GA-CG-SVM). All calculations were carried out by the LIBSVM freely available software (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). A total of 951 molecular descriptors in 13 different categories (e.g. constitutional, topological, geometrical, atom-centred fragments, connectivity indices, functional group counts, etc) were generated using the online program PCLIENT (http://www.vcclab.org/lab/pclient/). In the final GA–CG–SVM model, 25 structural descriptors were included. Five-fold cross-validation as well as independent external validation set including 76 compounds were used to validate the predictions. The obtained prediction accuracy confirmed by 5-fold cross-validation for the training set was 80% and for the independent test set was 86%, which is better than or comparable to the performances of other classification models in literature. The average cross-validated prediction accuracy for the positive (i.e. high bioavailability) compounds reached 99%, but the average accuracy for negative (i.e. low bioavailability) compounds was only 25%, indicating that that these could not be identified correctly. This result is consistent with other classification bioavailability models. For the independent validation set the cross-validated

accuracy of 97% for positive and 44% for negative compounds was reached. Although the modelling algorithm proposed by Ma et al. seems to be too complex to be transparent and easily reproducible, it is based on a rationale approach – bioavailability is determined by multiple factors and since GA–CG–SVM models cover a large range of molecular properties, they allow as many significant variables as possible to be captured. This study was performed for the largest and most diverse dataset to date. Thus, the proposed approach could be widely applicable to a broad range of compounds falling within various chemical classes.

The use of the Hologram Quantitative-Structure Property Relationship (HQSAR) approach to model human oral bioavailability was illustrated by Moda et al. (2007a). The authors compiled a large and diverse dataset of 302 compounds, falling within numerous chemical classes (enzyme inhibitors, receptor agonists and antagonists, antibiotics, analgesics, antivirals, anticancers, antibacterials, antifungals, antidepressants, antiepileptics, antihypertensives, anti-inflammatories, antiparasitics, anxiolytics, antipsychotics and antispasmodics). Groups of low (• 40%, 74 compounds), intermediate (41–80%, 127 compounds) and high (> 80%, 101 compounds) bioavailability compounds were distinguished according to the bioavailability values. A total of 250 training compounds and 52 test compounds were selected. HQSAR modeling was carried out using SYBYL 7.2. The most significant HQSAR model, based on molecular "fragment distinctions" (atoms (A), bonds (B), connections (C), hydrogen atoms (H), chirality (Ch), and donor and acceptor atoms (DA)) yielded the good statistics (cross-validated correlation coefficient $q^2 = 0.70$ and non cross-validated correlation coefficient $r^2 = 0.93$). The use of an external test set containing 52 molecules revealed a good predictive ability for compounds not included in the training set ($r^2 = 0.85$). Although the authors acknowledged some shortcomings of their approach (the model would not work well for the molecules which do not follow Lipinski's Rule of 5 or either are poorly soluble or not orally bioavailable), the predictive performance of the model was satisfactory both in terms of the coverage of chemical diversity and number of investigated compounds are concerned. Hence, it might be useful for human oral bioavailability estimation for structurally-diverse classes of chemicals.

A so-called Quantitative Structure-Bioavailability Relationship model, proposed by Andrews et al. (2000), led to some useful general conclusions about the effects of molecular substructures on bioavailability. The model was developed from a data set comprising 591 compounds from Glaxo Wellcome's internal database. For model development, the stepwise MLR procedure was applied to simple one-dimensional descriptors (including 608 substructure counts). The final MLR model, based on 85 descriptors, was internally validated by leave-one-out and leave-many-out cross validation (the regression coefficients were equal 0.63 and 0.58, respectively). The results suggested that some substructures (e.g. hydrogen bond donors, tetrazole, 4-aminopyridine, benzoquinone, dihydropyran, cyclohexanone, interior amino acid residues, aromatic and aliphatic ketones) are indicators of reduced bioavailability, whereas other substructures (hydrogen bond acceptors, halogens, N-terminal amino acid residues, aromatic and aliphatic esters) are indicators of increased bioavailability.

Yoshida and Topliss (2000) applied the Fuzzy Adaptive Least Squares FALS approach and the simplex technique (ORMUCS) to categorize a data set of 272 drugs. All (232) compounds in the training set were divided into four bioavailability classes (• 20%, 20-49%, 50-79% and • 80%). On the basis of physicochemical and structural factors, discriminant functions were developed to separate particular classes. Lipophilicity, expressed as the distribution coefficient at pH 6.5, was found to be a significant factor influencing bioavailability. The observation that acids generally had better bioavailability characteristics than bases (with neutral compounds between), led to the formulation of a new parameter, • log D (log $D_{6.5}$-log $D_{7.4}$), which proved to be an important contributor in improving the classification results. The

addition of 15 structural descriptors relating primarily to well-known metabolic processes yielded a QSAR equation which had a correct overall classification rate of 71% (97% within one class) and a Spearman rank correlation coefficient (*R*s) of 0.85. Nevertheless, the more accurate predictions were found for the compounds falling within higher bioavailability classes. The predictions for low bioavailability compounds were poor, probably due to their extensive enzymatic metabolism, which is difficult to model. The predictive power of the model was evaluated using a separate test set of 40 compounds, of which 60% (95% within one class) were correctly classified. The relationship formulated in this study identified significant factors influencing bioavailability and assigned them quantitative values expressing their contribution. This study may be potentially useful for further development of human oral bioavailability models as it identifies the set of significant descriptors.

One of the earliest studies on human oral bioavailability was performed by Hirono et al. (1994) and also was based on Fuzzy Adaptive Least Squares (FALS) methodology. The authors built a model on the basis of 188 drug molecules and classified all the compounds into three structure groups: A (having no aromatic ring), B (having aromatic hydrocarbon rings but no heteroaromatic rings) and C (having heteroaromatic rings). A separate model was built for each group on the basis of logP, MW and a range of discrete indicator variables relating to the presence of specific functional groups. The models were validated by leave-one-out cross validation. The results gave some insights about the effects of specific substructures. For example, for compounds in group A, •,•-saturated oxygen atoms contributed to the enhancement of bioavailability, whereas hydroxyl groups (liable to biotransformation in the gut wall and liver) reduced it. The study of Hirono et al. (1994) still might be potentially useful since it identifies structural features that could be used to develop human oral bioavailability models for a wider range of compounds.

6.3. Literature models for blood/brain barrier permeability

Literature models for blood/brain barrier permeability are summarised in Table 9.

A recent study by Obrezanova et al. (2008) illustrates a machine learning approach for building non-linear QSARs for logBB prediction. They applied a methodology called "Gaussian Processes", which is based on a Bayesian probabilistic approach and implemented in the BioFocus DPI's AutoModeler^{TM} software (http://www.biofocus.com/). Their model, based on 292 diverse compounds and 167 2D descriptors had a good predictive performance (r^2 = 0.74 for the validation set and 0.73 for the test set, and the overall Root Mean Squared Error RMSE = 0.34). The approach is presented by the authors as an automatic model building process for non-specialists. A potential disadvantage however is that the method generates models which are difficult to interpret and it is difficult to extract the contribution of each descriptor to the observed activity or property.

The Linear Free Energy Relationship approach (also used to model human intestinal absorption) has also been applied to blood/brain permeability prediction by Platts et al. (2001). By applying Multivariate Linear Regression (MLR) to a dataset of 148 diverse compounds (mainly drugs), they obtained a transparent QSAR incorporating 5 Abraham descriptors and an indicator variable (equal 1 for carboxylic acids and 0 for other compounds). The model was reported to show good statistics (R^2 = 0.74, standard deviation error, s = 0.34, and cross-validated correlation coefficient, R_{CV}^2 = 0.71). The study of Platts et al. (2001) is useful since the model algorithm is transparent and the coefficients in the model equation indicate the trends among descriptors significant for BBB permeation. The increasing size of molecules strongly enhances brain uptake, while increasing polarity/polarisability, hydrogen-bond acidity, basicity and the presence of carboxylic acid

groups have a detrimental influence. Moreover, the computation of logBB using Platts' method of determining descriptors (fragmentation scheme) is very fast and has been implemented in the commercially available ADME Boxes software (previously Pharma Algorithms; now ACD Labs, http://www.acdlabs.com/).

The use of the VolSurf technique, which is based on molecular interaction fields, has been used for blood/brain partitioning modelling, as demonstrated by Crivori et al. (2000). They built a binary decision model suitable for organic compounds on the basis of 230 diverse compounds and more than 70 VolSurf descriptors. The model was reported to have a prediction accuracy (assessed against an external test set) of 90% for BBB permeable molecules, and of 60% for non-permeable ones. Although a shortcoming of this approach is that it does not provide a transparent physical interpretation, due to the large number of descriptors, the computational procedure is fully automated and fast. Moreover, VolSurf approach-based models reach high levels of predictive accuracy. Hence, the methodology proposed by Cruciani et al. (2000a) is a valuable tool for the virtual screening of large datasets of diverse molecules.

6.4. Literature models for plasma protein binding

Literature models for plasma protein binding are summarised in Table 10.

Among the models proposed for human serum albumin (HSA) binding, the one that was based on the largest dataset of diverse molecules was proposed by Hajduk et al. (2003). By applying the group contribution methodology to an experimental dataset of dissociation constants for 889 compounds, they built a model for predicting the logarithm of the dissociation constant on the bias of 74 chemical groups ($logK = \bullet\, w_i x_i$, where w_i is the weighting coefficient of structural descriptor representing a particular chemical group and x_i is the number of times it appears in the molecule). The model showed good agreement with the experimental data (R^2 = 0.94, average error = 0.11). The good predictivity of model was confirmed by leave-group-out cross validation and randomization tests (Q^2 = 0.90). The analysis of structural descriptors and their weighting coefficients allowed the authors to find some general trends regarding the affinity of diverse molecules to HSA binding. Positively charged groups (e.g. cyclic and acyclic amines) having positive w_i values decreased HSA binding. Conversely, negatively charged groups (e.g. carboxylic and sulfonic acids), five- and six- membered rings as well as chlorine and fluorine substituents improved binding. The authors highlighted the general tendency of hydrophobic chemical substructures to enhance the process of HSA binding. The model reported by Hajduk et al. (2003) is of practical value in that it can be used for fast recognition of structural features making a molecule prone to HSA binding in large sets of diverse compounds. The disadvantage of the proposed model is that it cannot predict the dissociation constants for compounds possessing groups not represented among 74 used descriptors.

Several models have been proposed for predicting the whole plasma protein binding, usually expressed as percentage of PPB. A global model of human serum protein binding (% bound) was developed by Votano et al. (2006) on the basis of 1008 experimental values (808 training molecules and 200 test molecules) of human serum protein binding selected from publicly available sources. Using an initial set of 181 descriptors, four modelling techniques were applied to produce models: Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), k-Nearest Neighbours (kNN), and Support Vector Machines (SVM). With the exception of the MLR model, the ANN, kNN, and SVM QSARs were ensemble models. The final models included from 29 (kNN), 30 (MLR) and 33 (ANN) to 61 (SVM) descriptors. Training set correlation coefficients and mean absolute errors ranged from R^2 = 0.90 and

MAE = 7.6 for ANN to R^2 = 0.61 and MAE = 16.2 for MLR. Prediction results from the validation set yielded cross-validated (leave-one-out/10-fold cross validation) correlation coefficients and mean absolute errors which ranged from R^2 = 0.70 and MAE = 14.1 for ANN to R^2 = 0.59 and MAE = 18.3 for the SVM model. The authors performed a ranking of the descriptors and selected the 20 most significant ones. They concluded that, as far as human PPB modelling is concerned, the best results may be obtained by combining topological descriptors with logP. Moreover, on the basis of ANN and MLR models, they investigated the impact of structural features on protein binding processes. This analysis revealed that hydrophilicity, aromaticity, the presence of a ring structure, and the presence and bonding state of amines play important stereochemical roles in serum plasma protein binding of the compounds. Although ANN modelling requires some expertise and provides complex and difficult-to-interpret models, it does allow non-linear relationships between variables to be modelled, which may be particularly useful for heterogeneous datasets This study is valuable for further research in that it identifies interpretable descriptors and structural features relevant to logBB prediction, and the dataset on human serum protein binding compiled by the authors is available for further investigations.

6.5. Literature models for metabolism

A number of QSARs have been proposed in the literature for the modelling of metabolism (Table 4.11). These have generally been developed for phase I metabolism, and can be grouped into a number of categories according to the nature of their prediction: CYP substrate specificity, and CYP inhibition, and CYP regioselectivity.

In relation to QSAR models for CYP450 substrate recognition, Terfloth et al. (2007) investigated the application of several model-building techniques, namely: k-Nearest Neighbours (k-NN), C4.5/J48 decision tree, Multilayer Perceptron Neural Networks (MLP-NN), Radial Basis Function Neural Networks (RBF-NN), Logistic Regression (LR) and Support Vector Machine (SVM), to predict the isoform specificity for CYP450 3A4, 2D6 and 2C9 substrates. The authors used a dataset originally compiled by Manga et al. (2005), containing drugs which are predominately metabolised by CYP3A4, CYP2D6 or CYP2C9. The dataset to build and internally validate the model consisted of 146 (96 + 50) compounds (80 CYP3A4 substrates, 45 CYP2D6 substrates and 21 CYP2C9 substrates). For external validation purposes the authors selected 233 compounds (144 CYP3A4 substrates, 69 CYP2D6 substrates and 20 CYP2C9 substrates) from the Metabolite Database (http://www.symyx.com/). The applied descriptors included simple molecular properties and functional group accounts, topological descriptors, descriptors related to the shape of molecules or the distribution of interatomic distances considering the 3D structures of the molecules. The developed models were internally checked by the means of cross-validation. The (9-descriptor) model with the best results was established by combining automatic variable selection with the SVM technique. In leave-one-out cross-validation, 89% of all compounds in the training set were correctly classified. The achieved predictivity for an external data set of 233 compounds was equal 83%, a substantial improvement compared with the value achieved by Manga et al. (2005). Promising results were also obtained for the decision tree based model, which used three descriptors only and gave the values of 88% in LOO-CV and about 80% for the external data set. As this model consists of a simple decision tree having six branches and four leaves for the training set, it can be easily interpreted and used. It has been implemented in an on-line service to predict the isoform specificity (http://www.molecular-networks.com/online_demos/cyp450/).

An example of a model for CYP inhibition was proposed by Burton et al. (2006), who constructed classification models for human CYP1A2 and CYP2D6 inhibition using binary

decision trees. The modelling was based on 498 and 306 inhibition measures (IC50 and Ki values) for CYP1A2 and CYP2D6, respectively, collected from the Aureus Pharma (http://www.aureus-pharma.com/). To assess the performance of models, two external test sets of 34 and 58 molecules related to the CYP2D6 and CYP1A2 were used. The most predictive models were developed on the basis of Ki values. The decision tree for CYP2D6 had a goodness-of-fit to the training set of 90% (sensitivity 88%, specificity 92% and positive predictivity 90%). The external validation also was successful, as only two false negatives and two false positives were identified. The classification parameters were close to those obtained with the training set, namely the accuracy was 89%, sensitivity 91%, specificity 92% and precision 90%. For CYP1A2, the Ki-based model was characterised by the following goodness-of-fit statistics: accuracy = 89%, sensitivity = 95%, specificity = 83% and precision = 85%. The test set was significantly dissimilar to the training set, but the obtained statistics still were reasonable, i.e. 81% for accuracy, 76% for sensitivity, 86% for specificity and 85% for precision. The study of Burton et al. reveals several findings useful for further research. The authors demonstrated significant differences with models built with Ki or IC50 and concluded that the Ki values, defining the affinity of the inhibitor for the enzyme independently from the type/concentration of substrate and incubation conditions, are preferable to IC50 values or percentage of inhibition. The authors also identified a range of useful descriptors. The use of van der Waals surface area (VSA) descriptors was particularly efficient and allowed to develop models reaching 95% correct classification. 3D descriptors also provided promising results. An advantage of this approach is that the resulting decision tree models are easy and rapid to implement, as well as transparent and readily transferable.

In general models for regio-selectively, i.e. for predicting sites of metabolism on the substrate, have been based on mechanistic information relating to the substrate-enzyme interaction. An alternative approach, entirely QSAR-based and not simulating specific mechanisms and/or using explicit models of active sites, was proposed by Sheridan et al. (2007) for predicting CYP450 (3A4, 2D6, 2C9) sites of the metabolism The model was based on a dataset consisting of 532 diverse molecules (316 for CYP3A4, 124 for CYP2D6, and 92 for CYP2C9), collected mainly from the literature, but including also some proprietary in-house data. An external test set of 25 compounds was also used. In order to identify the most commonly observed potential sites of oxidation (sp3 carbons, sp2 carbons, sulphurs, etc.), the authors applied Random Forest (RF) technique to the dataset, using descriptors that describe the environment around each non-hydrogen atom in each molecule. In order to internally validate the model, both cross validation and external validation procedures were used. The predictions were compared with those obtained with earlier mechanistic model of Singh et al. (2003) and results using MetaSite software (Molecular Discovery) of Cruciani et al. (2005). As far as CYP3A4 was concerned, for the same set of 316 compounds, the oxidation site was in the top two atoms in 77%, 51% and 62% of the molecules for the QSAR model developed by Sheridan et al., the model of Singh et al. and MetaSite, respectively. This indicated that the final model proposed by Sheridan et al. for CYP3A4 was better than the other two. For CYP2D6 and CYP2C9 the predictions of Sheridan's model were only slightly better. This study is novel in that it shows that an empirical model for predicting regio-selectivity, developed on the basis of descriptors intrinsic to the candidate substrates and including no information about the active sites of the CYPs, can (in some cases) give results at least as good or better than the mechanism-based methods which simulate the chemical steps involved in the oxidation reaction. Furthermore, the authors identified several descriptors positively and negatively related to the oxidation sites of molecules, which is useful for further research and model development. Nevertheless, mechanistic approaches such as MetaSite still have significant advantages. MetaSite makes predictions based on the lability of hydrogens and orientation effects derived from the 3D structure of a CYP active site, independently of the

availability of pre-existing data. MetaSite can handle 3A4, 2D6, 2C9, 1A2, 2C9, and 2C19 and can be extended to any CYP for which a homology model can be generated. It is advantageous for enzymes such as CYP1A2 and CYP2C19, where there are not currently enough data in the literature to generate a QSAR model. Moreover, the MetaSite methodology is easy to use, fast and fully automated.

6.6. Literature models for excretion

Literature models for excretion are summarised in Table 12.

A series of so-called Quantitative Structure-Pharmacokinetic Relationship (QSPkR) models for the prediction of total (biliary and renal) clearance, CL_{tot}, were reported by Yap et al. (2006). The authors applied three statistical learning methods to a dataset of 503 various drugs, namely General Regression Neural Networks (GRNN), Support Vector Regression (SVR) and k-Nearest Neighbour (KNN), to explore their usefulness for building a global QSPkR model. Six different sets of molecular descriptors were evaluated for their usefulness in prediction.

The developed GRNN, SVR and KNN models were compared with a PLS linear model. The best performance was observed for SVR and GRNN models, which were characterised by 74.3% and 69.5% of compounds with the predicted CL_{tot} within two-fold error of the actual CL_{tot}, respectively. The best prediction accuracies were obtained for models which were based on a combination of different types of structural descriptors (constitutional, geometrical, topological and electrotopological ones). The authors also performed consensus modelling by combining the SVR and GRNN models into a single cQSPkR model, which slightly improved the modelling results. The study of Yap et al. indicates that statistical learning methods such as GRNN and SVR are able to capture a variety of multiple and interacting mechanisms involved in determining CL_{tot} better than linear models. Another useful conclusion is that a collection of various types of descriptors is more relevant for modeling the total clearance than individual specialised sets of descriptors, which are likely to neglect some important features. The most significant molecular properties found to influence the clearance of a compound were charge, molecular solvation, molecular size and flexibility. This study provides a useful basis for further research. However, the models would only be of practical value to the non-specialised user if encoded into a software tool.

For the prediction of human renal clearance, Doddareddy et al. (2006) developed models on the basis of 150 diverse CNS and non-CNS drugs, divided into training and test sets of 130 and 20 compounds, respectively. The authors utilized the VolSurf approach to explore the effect of VolSurf descriptors on renal clearance. For comparative purpose they also examined the usefulness of Molconn-Z topological descriptors. The authors performed Partial Least Squares (PLS) to develop renal clearance models on the basis of both VolSurf and Molconn-Z descriptors. The use of VolSurf descriptors resulted in a four-component model with the following statistics: cross-validated correlation coefficient $r^2 = 0.77$, standard deviation of error of predictions SDEP = 13.43 and standard deviation of error of calculations SDEC = 11.02. The use of Molconn-Z descriptors resulted in a PLS model with worse statistics ($r^2 = 0.53$, SDEP = 19.47, SDEC = 15.05). The authors also used two classification methods to divide the compounds into of low- and high-clearance groups. This led to the conclusion that both PLS models (VolSurf and Molconn-Z) were able to correctly predict the training set compounds in their respective groups of low (< 20%) and high (> 20%) renal clearance (80%-88% of the training compounds were predicted correctly, depending on the classification method). In the case of the test set compounds, one classification method (SIMCA) showed that the predictivity was better when Molconn-Z descriptors were used (85% of compounds

were predicted correctly) than when VolSurf descriptors were used (65% correct predictions). The other classification method (recursive partitioning) showed that about 70% of test compounds were correctly predicted by both models. The study of Doddareddy et al. is useful as it identifies an efficient approach for the quantitative and qualitative prediction of renal clearance. It also indicates that the Volsurf descriptors, which are based on 3D molecular fields, are more useful predictors of renal clearance than 2D topological descriptors. However, the most significant VolSurf descriptors were associated mainly with hydrophobicity/liphophilicity and the most important descriptor seemed to be the octanol/water partition coefficient logP, which is inversely proportional to the percentage renal clearance.

7. Conclusions

The progress made in the development of models for specific ADME properties, their potential usefulness and their shortcomings has been discussed in the subchapters above and summarised below. Overall, it can be concluded that a large number of QSARs and software tools have been developed, especially for the prediction of certain ADME properties (e.g. blood/brain barrier permeability, human intestinal absorption). However, their applicability in the dietary risk assessment of chemicals other than drugs is either poor or not established. This is a consequence of the fact that ADME models have been developed mainly for pharmaceutical purposes, and the available data sets are skewed toward drug molecules. If these models are applied to other classes of chemicals, the predictions may be unreliable, and in many cases the user will not be able to judge on this, since the applicability domains have not been explicitly defined and in many cases the training sets are confidential. Furthermore, in the case of software models, details of the predictive algorithm are not usually transparent. On the other hand, published studies have revealed a range of promising methodologies that can be used for further model development, and a number of easily-interpreted structural and physicochemical descriptors have been identified as useful predictors of ADME properties.

To promote the wider use of *in silico* models for ADME properties in the risk assessment of chemicals other than drugs, various significant research initiatives would need to be undertaken: a) it will be necessary to generate high-quality experimental datasets for classes of chemicals other than drugs (e.g. industrial chemicals, pollutants, food additives, pesticides); and b) the applicability of each model would have to be determined, on a case-by-case basis, by comparing its predictions with experimental data for chemical inventories of interest.

7.1. Conclusions regarding human intestinal absorption models

The majority of published models for human intestinal absorption have been developed using datasets including drugs and drug-like molecules, what creates a significant shortcoming as far as their applicability to different classes of chemicals is concerned. Furthermore, the published models are at the research stage, and not yet implemented into software suitable for the routine assessment of chemicals.

Nevertheless, some general findings have been identified that may be useful in further studies. The most significant descriptors for HIA are related to hydrogen bonding, molecular size, lipophilicity and surface polarity. Moreover, some generic functional groups which have detrimental impact on HIA have been identified, e.g. quaternary nitrogens and biphosphonates. The datasets used in modelling procedures should include, if possible, chemicals covering the whole range of %HIA values in order to avoid biases towards

poorly/highly absorbed compounds. Some compounds (usually those actively transported, insoluble or acting as P-glycoprotein substrates) appear as outliers or rule contradictors in HIA models – in such cases a set of preliminary models for active transport/solubility/P-gp binding could be developed before HIA prediction in order to identify the outliers and avoid final prediction errors. In addition, future research efforts should investigate ways of incorporating metabolic effects into QSAR models.

7.2. Conclusions regarding bioavailability models

Bioavailability is a very challenging property to model, due to the diversity of the underlying determinants, some of which (e.g. first-pass metabolism) are very difficult to model. One of the handicaps in bioavailability modelling is the paucity of data publicly available to the scientific community and the fact that the majority of the data available concerns mainly drugs and drug-like molecules.

Despite these difficulties, several attempts have been made to model human oral bioavailability, generally in categorical terms (e.g. high *vs* low bioavailability). These studies, summarised in Table 8, have resulted in a reasonable or good ability to identify high bioavailability compounds, but a relatively poor ability to identify low bioavailability compounds. Available studies also show that modelling strategies based on whole-molecule descriptors of diverse structures is not sufficient, as it does not allow to effectively characterise the first-pass metabolism. The more successful models employ well-defined substructures, which are probably related to different metabolism pathways.

7.3. Conclusions regarding Blood Brain Barrier models

There is a wealth of BBB permeability information published in the literature and available databases, which could potentially be applied by researchers to develop in silico models of brain penetration. However, the major shortcoming of existing data sets is that they tend to be relatively small (less than 100 compounds), they come from a variety of sources and may not be sufficiently consistent for modelling purposes. Other datasets were compiled specifically for drugs. Very few models have been proposed for determination of logBB for pollutants. Hence, one of the most urgent needs is the generation of larger and more diverse datasets with accurate measurements of logBB values. Nevertheless, the majority of recently developed QSAR models based on logBB data represent good predictivity as determined by both internal validation against the training set and external validation against test sets. There are a number of *in silico* models yielding logBB predictions of around 0.35-0.45 log units that could be used for screening purposes. By examining the wide variety of potentially useful molecular descriptors that have been reported, some important generalisations for further modelling studies can be made. Generally it is possible to distinguish two categories of descriptors. The first includes descriptors of size (i.e. molar refraction, connectivity and topological indices, molecular mass, surface area) while the second includes descriptors of polarity (i.e. polar surface area, partial charges, functions of hydrogen bond acid or hydrogen bond base groups). The descriptors from the first class are important predictors for the partitioning of non-polar compounds in the brain, whereas the descriptors from the second category express the features of polar molecules which are determine their tendency to partition in the blood.

7.4. Conclusions regarding models for plasma protein binding

The relatively small number of studies performed for plasma protein binding is a result of complexity of factors influencing the binding process on the one hand and the paucity of PPB

human data on the other. Large differences between data obtained from various species put into question the utility of models developed on non-human plasma proteins to predict human plasma protein binding. The majority of available human PPB models are based on data for drug molecules and tend to have a local character with applicability domains limited to small sets of structurally similar molecules. Although such models are relatively simple (they are based on relatively small number of descriptors, with lipophilicity being the most significant one) and probably easily reproducible and transferable, they cannot be applied to sets of structurally diverse compounds. However, a few investigations (discussed above) were based on broader datasets. Based on these studies, it can be concluded that lipophilicity alone is important but not sufficient to model PPB processes, especially in the case of large and diverse datasets concerned. It is necessary to use additional descriptors of various types (e.g. structural, topological, quantum mechanical) to obtain more complex and reliable human PPB models, and the use of non-linear modelling techniques may also be necessary. However, this is usually connected with a decreased transparency and reproducibility of the models.

7.5. Conclusions regarding models for metabolism

The utility of conventional QSARs predicting the metabolic fate of chemicals is highly limited. However, computer-based expert systems (COMPACT, META, MetabolExpert, METEOR, TIMES; see Table 5) have a much broader applicability.

A few QSAR models in the literature have provided some promising results for further research studies (discussed above). Most of these were designed to predict the phase I metabolism, with CYP450 isoforms playing a predominant role in the biotransformation of human drugs and xenobiotics. The modelling of phase II metabolism has not received as much attention; in most cases, these models have been developed for GST-catalyzed biotransformation.

Although progress is being made in the development of QSARs for metabolism, currently available models are typically derived from small data sets (only few of them are based on more than 100 compounds) and thus show poor predictivity for heterogenous sets of compounds. Most of the available QSARs have been developed for the purposes of drugs discovery and development. Furthermore, the model-building methodology, underlying training sets and model algorithms are often not transparent, which is an impediment to interpretation and reproducibility. A major bottleneck is the paucity of high quality and relevant experimental (*in vitro* or *in vivo*) data for use in model building and validation. Thus, it is difficult to make clear recommendations about which currently available literature models could be used in the dietary risk assessment of chemicals other than drugs. To make progress in this respect, more transparent descriptions of the applied approaches and training datasets are needed.

As far as modelling of CYP inhibition is concerned, literature QSARs are at an early stage of development as they usually give poor predictions when tested on the external sets of compounds. Much better results can be obtained from the models predicting the site of the metabolism (predictivity of 80% or more). The most challenging task seems to be modelling the rates of metabolism.

Significant improvement could probably be obtained by combining multiple *in silico* models for metabolism prediction (consensus modelling) along with physiologically based pharmacokinetic (PBPK) modelling utilising the data from different sources (*in silico*, *in vivo* and *in vitro*). However, this represents a long-term research effort.

7.6. Conclusions regarding excretion (clearance) models

The complexity of excretion processes and paucity of experimental data have hindered the development of models for excretion. Some efforts to model human total, urinary and (to a lesser extent) biliary clearance have been made only recently. These studies have identified some important trends governing the clearance processes, which form a useful basis for further research and model development. Most of the models are based on non-linear relationships and utilize large numbers of molecular descriptors in order to capture the multiple features affecting the clearance process. These models tend to be less transparently documented and thus of low reproducibility. However, if encoded into software tools, they could be practically useful. From the available literature, it seems that the software-based VolSurf approach, shown to be successful for modelling human intestinal absorption, oral bioavailability and blood/brain barrier penetration modelling, also works well for renal clearance prediction. Given the emphasis of published studies on drugs, the applicability of these approaches to other types of chemicals would require further investigation.

8. References

Abraham MH, Chadha HS & Mitchell RC (1995). Hydrogen-bonding. 36. Determination of blood brain distribution using octanol-water partition coefficients. Drug Design and Discovery 13(2), 123-131.

Abraham MH (2004). The factors that influence permeation across the blood-brain barrier. European Journal of Medicinal Chemistry 39(3), 235-240.

Abraham MH, Ibrahim A, Zhao Y & Acree WE Jr (2006). A data base for partition of volatile organic compounds and drugs from blood/plasma/serum to brain, and an LFER analysis of the data. Journal of Pharmaceutical Sciences 95(10), 2091-2100.

Abraham MH, Takács-Novák K & Mitchell RC (1997). On the partition of ampholytes: Application to blood-Brain distribution. Journal of Pharmaceutical Sciences 86(3), 310-315.

Abraham MH, Zhao YH, Le J, Hersey A, Luscombe CN, Reynolds DP, Beck G, Sherborne B & Cooper I (2002). On the mechanism of human intestinal absorption. European Journal of Medicinal Chemistry 37(7), 595-605.

Abraham MHC, Harpreet S, Mitchell RC (1994). Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. Journal of Pharmaceutical Sciences 83(9), 1257-1268.

Adenot M & Lahana R (2004). Blood-Brain Barrier Permeation Models: Discriminating between Potential CNS and Non-CNS Drugs Including P-Glycoprotein Substrates. Journal of Chemical Information and Computer Sciences 44(1), 239-248.

Afzelius L, Masimirembwa CM, Karlén A, Andersson TB & Zamora I (2002). Discriminant and quantitative PLS analysis of competitive CYP2C9 inhibitors versus non-inhibitors using alignment independent GRIND descriptors. Journal of Computer-Aided Molecular Design 16(7), 443-458.

Afzelius L, Zamora I, Masimirembwa CM, Karlen A, Andersson TB, Mecucci S, Baroni M & Cruciani G (2004). Conformer- and Alignment-Independent Model for Predicting Structurally Diverse Competitive CYP2C9 Inhibitors. Journal of Medicinal Chemistry 47(4), 907-914.

Afzelius L, Zamora I, Ridderström M, Andersson TB, Karlén A & Masimirembwa CM (2001). Competitive CYP2C9 Inhibitors: Enzyme Inhibition Studies, Protein Homology Modeling, and Three-Dimensional Quantitative Structure-Activity Relationship Analysis. Molecular Pharmacology 59(4), 909-919.

Agatonovic-Kustrin S, Beresford R & Yusof APM (2001). Theoretically-derived molecular descriptors important in human intestinal absorption. Journal of Pharmaceutical and Biomedical Analysis 25(2), 227-237.

Agatonovic-Kustrin S, Ling LH, Tham SY & Alany RG (2002). Molecular descriptors that influence the amount of drugs transfer into human breast milk. Journal of Pharmaceutical and Biomedical Analysis 29(1-2), 103-119.

Ajay, Bemis GW & Murcko MA (1999). Designing Libraries with CNS Activity. Journal of Medicinal Chemistry 42(24), 4942-4951.

Al-Fahemi JHA, Cooper DL & Allan NL (2007). Investigating the utility of momentum-space descriptors for predicting blood-brain barrier penetration. Journal of Molecular Graphics and Modelling 26(3), 607-612.

Allen DD & Geldenhuys WJ (2006). Molecular modeling of blood-brain barrier nutrient transporters: In silico basis for evaluation of potential drug delivery to the central nervous system. Life Sciences 78(10), 1029-1033.

Anderson RP, Butt TJ & Chadwick VS (1992). Hepatobiliary excretion of bacterial formyl-methionyl peptides in rat. Digestive Diseases and Sciences 37(2), 248-256.

Andrews CW, Bennett L & Yu L (2000). Predicting Human Oral Bioavailability of a Compound: Development of a Novel Quantitative Structure-Bioavailability Relationship. Pharmaceutical Research 17(6), 639-644.

Arimoto R, Prasad MA & Gifford EM (2005). Development of CYP3A4 Inhibition Models: Comparisons of Machine-Learning Techniques and Molecular Descriptors. Journal of Biomolecular Screening 10, 197-205.

Asikainen A, Tarhanen J, Poso A, Pasanen M, Alhava E & Juvonen RO (2003). Predictive value of comparative molecular field analysis modelling of naphthalene inhibition of human CYP2A6 and mouse CYP2A5 enzymes. Toxicology in Vitro 17(4), 449-455.

Atkinson F, Cole S, Green C & van de Waterbeemd H (2002). Lipophilicity and Other Parameters Affecting Brain Penetration. Current Medicinal Chemistry - Central Nervous System Agents 2, 229-240.

Bai JPF, Utis A, Crippen G, He H-D, Fischer V, Tullman R, Yin H-Q, Hsu C-P, Jiang L & Hwang K-K (2004). Use of Classification Regression Tree in Predicting Oral Absorption in Humans. Journal of Chemical Information and Computer Sciences 44(6), 2061-2069.

Bains W, Gilbert R, Sviridenko L, Gascon J, Scoffin R, Birchall K, Harvey I & Caldwell J (2002). Evolutionary computational methods to predict oral bioavailability QSPRs. Current Opinion in Drug Discovery & Development 5(1), 44-51.

Balakin KV, Ekins S, Bugrim A, Ivanenkov YA, Korolev D, Nikolsky YV, Ivashchenko AA, Savchuk NP & Nikolskaya T (2004b). Quantitative Structure-Metabolism Relationship Modeling Of Metabolic N-Dealkylation Reaction Rates. Drug Metabolism and Disposition 32(10), 1111-1120.

Balakin KV, Ekins S, Bugrim A, Ivanenkov YA, Korolev D, Nikolsky YV, Skorenko AV, Ivashchenko AA, Savchuk NP & Nikolskaya T (2004a). Kohonen maps for prediction of binding to human cytochrome P450 3A4. Drug Metabolism and Disposition 32(10), 1183-1189.

Bhutoria S & Ghoshal N (2008). A Novel Approach for the Identification of Selective Anticonvulsants Based on Differential Molecular Properties for TBPS Displacement and Anticonvulsant Activity: An Integrated QSAR Modelling. QSAR & Combinatorial Science 27(7), 876-889.

Brewster ME, Pop E, Huang M-J & Bodor N (1996). AM1-based model system for estimation of brain/blood concentration ratios. International Journal of Quantum Chemistry 60(8), 1775-1787.

Burton J, Ijjaali I, Barberan O, Petitet F, Vercauteren DP & Michel A (2006). Recursive Partitioning for the Prediction of Cytochromes P450 2D6 and 1A2 Inhibition: Importance of the Quality of the Dataset. Journal of Medicinal Chemistry 49(21), 6231-6240.

Cabrera MA, Bermejo M, Pérez M & Ramos R (2004). TOPS-MODE approach for the prediction of blood-brain barrier permeation. Journal of Pharmaceutical Sciences 93(7), 1701-1717.

Calder JG & Ganellin CR (1994). Predicting the brain-penetrating capability of histaminergic compounds. Drug Design and Discovery 11(4), 259-268.

Cantelli Forti G, Guerra MC, Barbaro AM, Hrelia P, Biagi GL & Borea PA (2002). Relationship between lipophilic character and urinary excretion of nitroimidazoles and nitrothiazoles in rats. Journal of Medicinal Chemistry 29(4), 555-561.

Chohan KK, Paine SW, Mistry J, Barton P & Davis AM (2005). A Rapid Computational Filter for Cytochrome P450 1A2 Inhibition Potential of Compound Libraries. Journal of Medicinal Chemistry 48(16), 5154-5161.

Chohan KK, Paine SW & Waters NJ (2006). Quantitative Structure Activity Relationships in Drug Metabolism. Current Topics in Medicinal Chemistry 6, 1569-1578.

Chohan KK, Paine SW & Waters NJ (2008). Advancements in predictive in silico models for ADME. Current Chemical Biology 2, 215-228.

Cianchetta G, Yi L, Singleton R, Zhang M,Wildgoose M, Rampe D, Kang J & Vaz RJ (2006). Molecular Interaction Fields in ADME and safety, in: Molecular Interaction Fields (G. Cruciani ed). Wiley-VCH, Weinheim, pp. 197-218.

Clark DE (1999a). Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. Journal of Pharmaceutical Sciences 88(8), 807-814.

Clark DE (1999b). Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. Journal of Pharmaceutical Sciences 88(8), 815-821.

Clark DE (2003). In silico prediction of blood-brain barrier permeation. Drug Discovery Today 8(20), 927-933.

Clark DE (2007). In silico ADMET tools: a dawn of a new generation? Expert Opinion on Drug Discovery 2, 1423-1429.

Clark DEG, & Peter DJ (2002). Progress in computational methods for the prediction of ADMET properties. Current Opinion in Drug Discovery & Development 5(3), 382-390.

Colmenarejo G (2003). In silico prediction of drug-binding strengths to human serum albumin. Medicinal Research Reviews 23(3), 275-301.

Colmenarejo G (2005). In Silico ADME Prediction: Data Sets and Models. Current Computer-Aided Drug Design 1, 365-376.

Colmenarejo G, Alvarez-Pedraglio A & Lavandera J-L (2001). Cheminformatic Models To Predict Binding Affinities to Human Serum Albumin. Journal of Medicinal Chemistry 44(25), 4370-4378.

Crivori P, Cruciani G, Carrupt P-A & Testa B (2000). Predicting Blood• Brain Barrier Permeation from Three-Dimensional Molecular Structure. Journal of Medicinal Chemistry 43(11), 2204-2216.

Crivori P & Poggesi I (2005). Predictive Model for Identifying Potential CYP2D6 Inhibitors. Basic & Clinical Pharmacology & Toxicology 96, 251-253.

Crivori P & Poggesi I (2006). Computational approaches for predicting CYP-related metabolism properties in the screening of new drugs. European Journal of Medicinal Chemistry 41(7), 795-808.

Crivori P, Zamora I, Speed B, Orrenius C & Poggesi I (2004). Model based on GRID-derived descriptors for estimating CYP3A4 enzyme stability of potential drug candidates. Journal of Computer-Aided Molecular Design 18(3), 155-166.

Cross S & Cruciani G (2010). Molecular fields in drug discovery: getting old or reaching maturity? Drug Discovery Today 15(1-2), 23-32.

Cruciani G, Carosati E, De Boeck B, Ethirajulu K, Mackie C, Howe T & Vianello R (2005). MetaSite: Understanding Metabolism in Human Cytochromes from the Perspective of the Chemist. Journal of Medicinal Chemistry 48(22), 6970-6979.

Cruciani G, Crivori P, Carrupt PA & Testa B (2000a). Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. Journal of Molecular Structure: THEOCHEM 503(1-2), 17-30.

Cruciani G, Pastor M & Guba W (2000b). VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. European Journal of Pharmaceutical Sciences 11(Supplement 2), S29-S39.

Cupid BC (1999). Quantitative structure-metabolism relationships (QSMR) using computational chemistry: pattern recognition analysis and statistical prediction of phase II conjugation reactions of substituted benzoic acids in the rat. Xenobiotica 29(1), 27-42.

Cupid BC, Beddell CR, Lindon JC, Wilson ID & Nicholson JK (1996). Quantitative structure-metabolism relationships for substituted benzoic acids in the rabbit: prediction of urinary excretion

of glycine and glucuronide conjugates. Xenobiotica: The fate and safety evaluation of foreign compounds in biological systems 26(2), 157 - 176.

Darvas F (1988). Predicting metabolic pathways by logic programming. Journal of Molecular Graphics 6(2), 80-86.

de Graaf C, Oostenbrink C, Keizers PHJ, van der Wijst T, Jongejan A & Vermeulen NPE (2006). Catalytic Site Prediction and Virtual Screening of Cytochrome P450 2D6 Substrates by Consideration of Water and Rescoring in Automated Docking. Journal of Medicinal Chemistry 49(8), 2417-2430.

de Graaf C, Vermeulen NPE & Feenstra KA (2005). Cytochrome P450 in Silico: An Integrative Modeling Approach. Journal of Medicinal Chemistry 48(8), 2725-2755.

de Groot MJ, Ackland MJ, Horne VA, Alex AA & Jones BC (1999a). Novel Approach To Predicting P450-Mediated Drug Metabolism: Development of a Combined Protein and Pharmacophore Model for CYP2D6. Journal of Medicinal Chemistry 42(9), 1515-1524.

de Groot MJ, Ackland MJ, Horne VA, Alex AA & Jones BC (1999b). Novel Approach To Predicting P450-Mediated Drug Metabolism: Development of a Combined Protein and Pharmacophore Model for CYP2D6. Journal of Medicinal Chemistry 42(9), 1515-1524.

de Groot MJ, Alex AA & Jones BC (2002). Development of a Combined Protein and Pharmacophore Model for Cytochrome P450 2C9. Journal of Medicinal Chemistry 45(10), 1983-1993.

de Groot MJ, Vermeulen NPE, Kramer JD, van Acker FAA & Donne-Op den Kelder GM (1996). A Three-Dimensional Protein Model for Human Cytochrome P450 2D6 Based on the Crystal Structures of P450 101, P450 102, and P450 108. Chemical Research in Toxicology 9(7), 1079-1091.

de Rienzo F, Fanelli F, Menziani MC & De Benedetti PG (2000). Theoretical investigation of substrate specificity for cytochromes P450 IA2, P450 IID6 and P450 IIIA4. Journal of Computer-Aided Molecular Design 14(1), 93-116.

Dearden J & Worth A (2007). In Silico Prediction of Physicochemical Properties. JRC Report EUR 23051 EN. http://ecb.jrc.ec.europa.eu/qsar/publications/

Dearden JC (2007). In silico prediction of ADMET properties: how far have we come? Expert Opinion on Drug Metabolism & Toxicology 3(5), 635-639.

Deconinck E, Hancock T, Coomans D, Massart DL & Heyden YV (2005). Classification of drugs in absorption classes using the classification and regression trees (CART) methodology. Journal of Pharmaceutical and Biomedical Analysis 39(1-2), 91-103.

Deconinck E, Zhang MH, Coomans D & Heyden YV (2007). Evaluation of boosted regression trees (BRTs) and two-step BRT procedures to model and predict blood-brain barrier passage. Journal of Chemometrics 21(7-9), 280-291.

Deconinck E, Zhang MH, Coomans D & Vander Heyden Y (2006). Classification Tree Models for the Prediction of Blood• Brain Barrier Passage of Drugs. Journal of Chemical Information and Modeling 46(3), 1410-1419.

Deeb O & Hemmateenejad B (2007). ANN-QSAR Model of Drug-binding to Human Serum Albumin. Chemical Biology & Drug Design 70(1), 19-29.

Delisle RK, Lowrie JF, Hobbs DW & Diller DJ (2005). Computational ADME/Tox modeling: aiding understanding and enhancing decision making in drug design. Current Computer-Aided Drug Design 1, 325-345.

Deretey E, Feher M & Schmidt JM (2002). Rapid Prediction of Human Intestinal Absorption. Quantitative Structure-Activity Relationships 21(5).

Doddareddy MR, Cho YS, Koh HY, Kim DH & Pae AN (2006). In Silico Renal Clearance Model Using Classical Volsurf Approach. Journal of Chemical Information and Modeling 46(3), 1312-1320.

Doniger S, Hofmann T & Yeh J (2002). Predicting CNS Permeability of Drug Molecules: Comparison of Neural Network and Support Vector Machine Algorithms. Journal of Computational Biology 9(6), 849-864.

Dorronsoro I, Chana A, Abasolo MI, Castro A, Gil C, Stud M & Martinez A (2004). CODES/neural network model: A useful tool for in silico prediction of oral absorption and blood-brain barrier permeability of structurally diverse drugs. QSAR & Combinatorial Science 23(2-3), 89-98.

Dureja H & Madan A (2007). Validation of topochemical models for the prediction of permeability through the blood-brain barrier. Acta Pharmaceutica 57(4), 451-467.

Dureja H & Madan AK (2006). Topochemical models for the prediction of permeability through blood-brain barrier. International Journal of Pharmaceutics 323(1-2), 27-33.

Egan WJ, Merz KM & Baldwin JJ (2000). Prediction of Drug Absorption Using Multivariate Statistics. Journal of Medicinal Chemistry 43(21), 3867-3877.

Ekins S, Andreyev S, Ryabov A, Kirillov E, Rakhmatulin EA, Bugrim A & Nikolskaya T (2005). Computational prediction of human drug metabolism. Expert Opinion on Drug Metabolism & Toxicology 1(2), 303-324.

Ekins S, Berbaum J & Harrison RK (2003). Generation and validation of rapid computational filters for CYP2D6 and CYP3A4. Drug Metabolism and Disposition 31(9), 1077-1080.

Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH & Wrighton SA (2000a). Three- and Four-Dimensional-Quantitative Structure Activity Relationship (3D/4D-QSAR) Analyses of CYP2C9 Inhibitors. Drug Metabolism and Disposition 28(8), 994-1002.

Ekins S, Gianpaolo B, Wikel JH, Wrighton SA (1999a). Three-Dimensional-Quantitative Structure Activity Relationship Analysis of Cytochrome P-450 3A4 Substrates. Journal of Pharmacology and Experimental Terapeutics 291, 424-433.

Ekins S, Gianpaolo B, Binkley S, Gillespie JS, Ring BJ, Wikel JH & Wrighton SA (1999b). Three- and Four-Dimensional Quantitative Structure Activity Relationship Analyses of Cytochrome P-450 3A4 Inhibitors. Journal Of Pharmacology And Experimental Therapeutics 290, 429-438.

Ekins S, Bravi G, Ring BJ, Gillespie TA, Gillespie JS, Vandenbranden M, Wrighton SA & Wikel JH (1999c). Three-Dimensional Quantitative Structure Activity Relationship Analyses of Substrates for CYP2B6. Journal of Pharmacology and Experimental Therapeutics 288(1), 21-29.

Ekins S, Nikolsky Y & Nikolskaya T (2005). Techniques: Application of systems biology to absorption, distribution, metabolism, excretion and toxicity. Trends in Pharmacological Sciences 26(4), 202-209.

Ekins S & Obach RS (2000). Three-Dimensional Quantitative Structure Activity Relationship Computational Approaches for Prediction of Human In Vitro Intrinsic Clearance. Journal of Pharmacology and Experimental Therapeutics 295(2), 463-473.

Ekins S & Williams AJ (2010). Precompetitive preclinical ADME/Tox data: set it free on the web to facilitate computational model building and assist drug development. Lab on a Chip.

Ekins S, Mestres J & Testa B (2007a). In silico pharmacology for drug discovery: methods for virtual ligand screening and profiling. British Journal of Pharmacology 152(1), 9-20.

Ekins S, Mestres J & Testa B (2007b). In silico pharmacology for drug discovery: applications to targets and beyond. British Journal of Pharmacology 152(1), 21-37.

Ekins S, Embrechts MJ, Breneman CM, Jim K & Wery J-P (2007c). Novel Applications of Kernel-Partial Least Squares to Modeling a Comprehensive Array of Properties for Drug Discovery. In Computational Toxicology (S Ekins, ed), pp. 403-432.

El-Deeb IM, Bayoumi SM, El-Sherbeny MA & Abdel-Aziz AAM (2010). Synthesis and antitumor evaluation of novel cyclic arylsulfonylureas: ADME-T and pharmacophore prediction. European Journal of Medicinal Chemistry 45(6), 2516-2530.

Embrechts MJ & Ekins S (2007). Classification of Metabolites with Kernel-Partial Least Squares (K-PLS). Drug Metabolism and Disposition 35(3), 325-327.

Ertl P, Rohde B & Selzer P (2000). Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. Journal of Medicinal Chemistry 43(20), 3714-3717.

Estrada E, Uriarte E, Molina E, Simon-Manso Y & Milne GWA (2006). An Integrated in Silico Analysis of Drug-Binding to Human Serum Albumin. Journal of Chemical Information and Modeling 46(6), 2709-2724.

Ethell BT, Ekins S, Wang J & Burchell B (2002). Quantitative Structure Activity Relationships for the Glucuronidation of Simple Phenols by Expressed Human UGT1A6 and UGT1A9. Drug Metabolism and Disposition 30(6), 734-738.

Ewing T & Feher M (2010). Forecasting CYP2D6 and CYP3A4 Risk with a Global/Local Fusion Model of CYP450 Inhibition. Molecular Informatics 29(1-2), 127-141.

Feher M, Sourial E & Schmidt JM (2000). A simple model for the prediction of blood-brain partitioning. International Journal of Pharmaceutics 201(2), 239-247.

Fox T & Kriegl JM (2006). Machine Learning Techniques for In Silico Modeling of Drug Metabolism. Current Topics in Medicinal Chemistry 6, 1579-1591.

Franklin RB (2009). In silico studies in ADME/Tox: caveat emptor. Current Computer-Aided Drug Design 5, 128-138. However beautiful the strategy, you should occasionally look at the results. Winston Churchill (1874-1965). British Statesman.

Fu XC, Wang GP, Liang WQ & Yu QS (2004). Predicting blood-brain barrier penetration of drugs using an artificial neural network. Pharmazie 59(2), 126-130.

Fu XW, G. P.; Gao, J. Q.; Zhan, S. Y.; Liang, W. Q. (2007). Prediction of plasma protein binding of cephalosporins using an artificial neural network Pharmazie 62(2), 157-158.

Garg P & Verma J (2006). In Silico Prediction of Blood Brain Barrier Permeability: An Artificial Neural Network Model. Journal of Chemical Information and Modeling 46(1), 289-297.

Geldenhuys WJ, Manda VK, Mittapalli RK, Van der Schyf CJ, Crooks PA, Dwoskin LP, Allen DD & Lockman PR (2010). Predictive screening model for potential vector-mediated transport of cationic substrates at the blood-brain barrier choline transporter. Bioorganic & Medicinal Chemistry Letters 20(3), 870-877.

Gerebtzoff G & Seelig A (2006). In Silico Prediction of Blood• Brain Barrier Permeation Using the Calculated Molecular Cross-Sectional Area as Main Parameter. Journal of Chemical Information and Modeling 46(6), 2638-2650.

Ghauri F, Blackledge CA, Glen RC, Sweatman BC, Lindon JC, Beddell CR, Wilson ID & JK. N (1992). Quantitative structure-metabolism relationships for substituted benzoic acids in the rat. Computational chemistry, NMR spectroscopy and pattern recognition studies. Biochemical Pharmacology 44, 1935-1946.

Ghuloum AM, Sage CR & Jain AN (1999). Molecular Hashkeys: A Novel Method for Molecular Characterization and Its Application for Predicting Important Pharmaceutical Properties of Molecules. Journal of Medicinal Chemistry 42(10), 1739-1748.

Gleeson MP (2007). Plasma Protein Binding Affinity and Its Relationship to Molecular Structure: An In-silico Analysis. Journal of Medicinal Chemistry 50(1), 101-112.

Gleeson MP (2008). Generation of a Set of Simple, Interpretable ADMET Rules of Thumb. Journal of Medicinal Chemistry 51(4), 817-834.

Gola J, Obrezanova O, Champness E & Segall M (2006). ADMET Property Prediction: The State of the Art and Current Challenges. QSAR & Combinatorial Science 25(12), 1172-1180.

Goodwin JT & Clark DE (2005). In Silico Predictions of Blood-Brain Barrier Penetration: Considerations to "Keep in Mind". Journal of Pharmacology and Experimental Therapeutics 315(2), 477-483.

Greene N, Judson P, Langowski J & Marchant C (1999). Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. SAR and QSAR in Environmental Research 10, 299-314.

Gunturi SB, Narayanan R & Khandelwal A (2006). In silico ADME modelling 2: Computational models to predict human serum albumin binding affinity using ant colony systems. Bioorganic & Medicinal Chemistry 14(12), 4118-4129.

Hajduk PJ, Mendoza R, Petros AM, Huth JR, Bures M, Fesik SW & Martin YC (2003). Ligand binding to domain-3 of human serum albumin: a chemometric analysis. Journal of Computer-Aided Molecular Design 17(2), 93-102.

Haji-Momenian S, Rieger JM, Macdonald TL & Brown ML (2003). Comparative molecular field analysis and QSAR on substrates binding to cytochrome P450 2D6. Bioorganic & Medicinal Chemistry 11(24), 5545-5554.

Hall LH, Hall MH, Kier LB, Parham M & Votano JR (2006). Interpretation of the role of the electrotopological state and molecular connectivity indices in the prediction of physical properties and ADME-Tox behavior case study: human plasma protein binding Solvay Pharmaceuticals Conferences 6: Virtual ADMET Assessment in Target Selection and Maturation, 67-99.

Hall LM, Hall LH & Kier LB (2003). QSAR modeling of • -lactam binding to human serum proteins. Journal of Computer-Aided Molecular Design 17(2), 103-118.

Hasegawa K, Koyama M & Funatsu K (2010). Quantitative Prediction of Regioselectivity Toward Cytochrome P450/3A4 Using Machine Learning Approaches. Molecular Informatics 29(3), 243-249.

Hirono S, Nakagome I, Hirano H, Yoshii F & Moriguchi I (1994). Non-congeneric structure-pharmacokinetic property correlation studies using fuzzy adaptive least-squares: volume of distribution. Biological & Pharmaceutical Bulletin 17(5), 686-690.

Hollosy F, Valko K, Hersey A, Nunhuck S, Keri G & Bevan C (2006). Estimation of Volume of Distribution in Humans from High Throughput HPLC-Based Measurements of Human Serum Albumin Binding and Immobilized Artificial Membrane Partitioning. Journal of Medicinal Chemistry 49(24), 6958-6971.

Holm R & Hoest J (2004). Successful in silico predicting of intestinal lymphatic transfer. International Journal of Pharmaceutics 272(1-2), 189-193.

Holmes E, Sweatman BC, Bollard ME, Blackledge CA, Beddell CR, Wilson ID, Lindon JC & Nicholson JK (1995). Prediction of urinary sulphate and glucuronide conjugate excretion for substituted phenols in the rat using quantitative structure-metabolism relationships. Xenobiotica 25(12), 1269-1281.

Hou T, Wang J, Zhang W, Wang W & Xu X (2006). Recent Advances in Computational Prediction of Drug Absorption and Permeability in Drug Discovery. Current Medicinal Chemistry 13, 2653-2667.

Hou T, Wang J, Zhang W & Xu X (2007a). ADME Evaluation in Drug Discovery. 6. Can Oral Bioavailability in Humans Be Effectively Predicted by Simple Molecular Property-Based Rules? Journal of Chemical Information and Modeling 47(2), 460-463.

Hou T, Wang J, Zhang W & Xu X (2007b). ADME Evaluation in Drug Discovery. 7. Prediction of Oral Absorption by Correlation and Classification. Journal of Chemical Information and Modeling 47(1), 208-218.

Hou T, Wang J & Li Y (2007c). ADME Evaluation in Drug Discovery. 8. The Prediction of Human Intestinal Absorption by a Support Vector Machine. Journal of Chemical Information and Modeling 47(6), 2408-2415.

Hou T & Wang J (2008). Structure – ADME relationship: still a long way to go? Expert Opinion on Drug Metabolism & Toxicology 4(6), 759-770.

Hou T & Xu X (2002). ADME evaluation in drug discovery. Journal of Molecular Modeling 8(12), 337-349.

Hou TJ & Xu XJ (2003). ADME Evaluation in Drug Discovery. 3. Modeling Blood-Brain Barrier Partitioning Using Simple Molecular Descriptors. Journal of Chemical Information and Computer Sciences 43(6), 2137-2152.

Hu Y, Unwalla R, Denny R, Bikker J, Di L & Humblet C (2010). Development of QSAR models for microsomal stability: identification of good and bad structural features for rat, human and mouse microsomal stability. Journal of Computer-Aided Molecular Design 24(1), 23-35.

Hutter MC (2003). Prediction of blood–brain barrier permeation using quantum chemically derived information. Journal of Computer-Aided Molecular Design 17(7), 415-443.

Hyland R, Jones B & van de Waterbeemd H (2006). Utility of human/human-derived reagents in drug discovery and development: An industrial perspective. Environmental Toxicology and Pharmacology 21(2), 179-183.

Iori F, Fonseca Rd, Ramos MJ & Menziani MC (2005). Theoretical quantitative structure-activity relationships of flavone ligands interacting with cytochrome P450 1A1 and 1A2 isozymes. Bioorganic & Medicinal Chemistry 13(14), 4366-4374.

Iyer M, Mishra R, Han Y & Hopfinger AJ (2002). Predicting Blood–Brain Barrier Partitioning of Organic Molecules Using Membrane–Interaction QSAR Analysis. Pharmaceutical Research 19(11), 1611-1621.

Jamei M, Marciniak S, Feng K, Barnett A, Tucker G & Rostami-Hodjegan A (2009). The Simcyp® Population-based ADME Simulator. Expert Opinion on Drug Metabolism & Toxicology 5(2), 211-223.

Jacobs MN, Janssens W, Bernauer U, Brandon E, Coecke S, Combes R, Edwards P, Freidig A, Freyberger A, Kolanczyk R, Mc Ardle C, Mekenyan O, Schmieder P, Schrader T, Takeyoshi M & Burg Bvd (2008). The Use of Metabolising Systems for In Vitro Testing of Endocrine Disruptors. Current Drug Metabolism 9, 796-826.

Jensen B, Sørensen M, Kissmeyer A-M, Björkling F, Sonne K, Engelsen S & Nørgaard L (2003). Prediction of in vitro metabolic stability of calcitriol analogs by QSAR. Journal of Computer-Aided Molecular Design 17(12), 849-859.

Jones JP, He M, Trager WF & Rettie AE (1996b). Three-dimensional quantitative structure-activity relationship for inhibitors of cytochrome P4502C9. Drug Metabolism and Disposition 24(1), 1-6.

Jones JP, Mysinger M & Korzekwa KR (2002). Computational Models for Cytochrome P450: A Predictive Electronic Model for Aromatic Oxidation and Hydrogen Atom Abstraction. Drug Metabolism and Disposition 30(1), 7-12.

Jones R, Connolly PC, Klamt A & Diedenhofen M (2005). Use of Surface Charges from DFT Calculations To Predict Intestinal Absorption. Journal of Chemical Information and Modeling 45(5), 1337-1342.

Jung E, Kim J, Kim M, Jung D, Rhee H, Shin J-M, Choi K, Kang S-K, Kim M-K, Yun C-H, Choi Y-J & Choi S-H (2007). Artificial neural network models for prediction of intestinal permeability of oligopeptides. BMC Bioinformatics 8(1), 245.

Kalgutkar AS, Gardner I, Obach RS, Shaffer CL, Callegari E, Henne KR, Mutlib AE, Dalvie DK, Lee JS, Nakai Y, O'Donnell JP, Boer J & Harriman SP (2005). A Comprehensive Listing of Bioactivation Pathways of Organic Functional Groups. Current Drug Metabolism 6, 161-225.

Kaliszan R & Markuszewski M (1996). Brain/blood distribution described by a combination of partition coefficient and molecular mass. International Journal of Pharmaceutics 145(1-2), 9-16.

Karalis V, Tsantili-Kakoulidou A & Macheras P (2002). Multivariate Statistics of Disposition Pharmacokinetic Parameters for Structurally Unrelated Drugs Used in Therapeutics. Pharmaceutical Research 19(12), 1827-1834.

Karalis V, Tsantili-Kakoulidou A & Macheras P (2003). Quantitative structure-pharmacokinetic relationships for disposition parameters of cephalosporins. European Journal of Pharmaceutical Sciences 20(1), 115-123.

Karelson M, Dobchev D, Tamm T, Tulp I, Janes J, Tamm K, Lomaka A, Savchenko D, Karelson G (2008). Correlation of blood-brain penetration and human serum albumin binding with theoretical descriptors ARKIVOC(16), 38-60.

Katritzky AR, Kuanar M, Slavov S, Dobchev DA, Fara DC, Karelson M, Acree Jr WE, Solov'ev VP & Varnek A (2006). Correlation of blood-brain penetration using structural descriptors. Bioorganic & Medicinal Chemistry 14(14), 4888-4917.

Kaznessis YN (2005). A Review of Methods for Computational Prediction of Blood-Brain Partitioning. Current Medicinal Chemistry - Central Nervous System Agents 5, 185-191.

Kaznessis YN, Snow ME & Blankley CJ (2001). Prediction of blood-brain partitioning using Monte Carlo simulations of molecules in water. Journal of Computer-Aided Molecular Design 15(8), 697-708.

Kelder J, Grootenhuis PDJ, Bayada DM, Delbressine LPC & Ploemen J-P (1999). Polar Molecular Surface as a Dominating Determinant for Oral Absorption and Brain Penetration of Drugs. Pharmaceutical Research 16(10), 1514-1519.

Kemp CA, Flanagan JU, van Eldik AJ, Marechal J-D, Wolf CR, Roberts GCK, Paine MJI & Sutcliffe MJ (2004). Validation of Model of Cytochrome P450 2D6: An in Silico Tool for Predicting Metabolism and Inhibition. Journal of Medicinal Chemistry 47(22), 5340-5346.

Keseru GM & Molnar L (2001). High-Throughput Prediction of Blood• Brain Partitioning: A Thermodynamic Approach. Journal of Chemical Information and Computer Sciences 41(1), 120-128.

Khan MTH & Sylte I (2007). Predictive QSAR Modeling for the Successful Predictions of the ADMET Properties of Candidate Drug Molecules. Current Drug Discovery Technologies 4, 141-149.

Kharkar PS (2010). Two-Dimensional (2D) In Silico Models for Absorption, Distribution, Metabolism, Excretion and Toxicity (ADME/T) in Drug Discovery. Current Topics in Medicinal Chemistry 10, 116-126.

Kim K (1991). Quantitative structure-activity relationships of the metabolism of drugs by uridine diphosphate glucuronosyltransferase Journal of Pharmaceutical Sciences 80, 966-970.

Klamt A, Eckert F & Hornig M (2001). COSMO-RS: A novel view to physiological solvation and partition questions. Journal of Computer-Aided Molecular Design 15(4), 355-365.

Klon AE, Lowrie JF & Diller DJ (2006). Improved Naive Bayesian Modeling of Numerical Data for Absorption, Distribution, Metabolism and Excretion (ADME) Property Prediction. ChemInform 37(49).

Klopman G, Dimayuga M & Talafous J (1994). META. 1. A Program for the Evaluation of Metabolic Transformation of Chemicals. Journal of Chemical Information and Computer Sciences 34(6), 1320-1325.

Klopman G, Stefan LR & Saiakhov RD (2002). ADME evaluation: 2. A computer model for the prediction of intestinal absorption in humans. European Journal of Pharmaceutical Sciences 17(4-5), 253-263.

Klopman G, Tu M & Fan BT (1999). META 4. Prediction of the metabolism of polycyclic aromatic hydrocarbons. Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta) 102(1), 33-38.

Klopman G, Tu M & Talafous J (1997). META. 3. A Genetic Algorithm for Metabolic Transform Priorities Optimization. Journal of Chemical Information and Computer Sciences 37(2), 329-334.

Konovalov DA, Coomans D, Deconinck E & Vander Heyden Y (2007). Benchmarking of QSAR Models for Blood-Brain Barrier Permeation. Journal of Chemical Information and Modeling 47(4), 1648-1656.

Konovalov DA, Sim N, Deconinck E, Vander Heyden Y & Coomans D (2008). Statistical Confidence for Variable Selection in QSAR Models via Monte Carlo Cross-Validation. Journal of Chemical Information and Modeling 48(2), 370-383.

Korhonen LE, Rahnasto M, Mahonen NJ, Wittekindt C, Poso A, Juvonen RO & Raunio H (2005). Predictive Three-Dimensional Quantitative Structure• Activity Relationship of Cytochrome P450 1A2 Inhibitors. Journal of Medicinal Chemistry 48(11), 3808-3815.

Korolev D, Balakin KV, Nikolsky Y, Kirillov E, Ivanenkov YA, Savchuk NP, Ivashchenko AA & Nikolskaya T (2003). Modeling of Human Cytochrome P450-Mediated Drug Metabolism Using Unsupervised Machine Learning Approach. Journal of Medicinal Chemistry 46(17), 3631-3643.

Kortagere S, Ekins S (2010). Troubleshooting computational methods in drug discovery. Journal of Pharmacological and Toxicological Methods 61(2), 67-75.

Korzekwa KR, Jones JP & Gillette JR (1990). Theoretical studies on cytochrome P-450 mediated hydroxylation: a predictive model for hydrogen atom abstractions. Journal of the American Chemical Society 112(19), 7042-7046.

Koymans L, Vermeulen NPE, Van Acker SABE, Te Koppele JM, Heykants JJP, Lavrijsen K, Meuldermans W & Donne-Op den Kelder GM (1992). A predictive model for substrates of cytochrome P450-debrisoquine (2D6). Chemical Research in Toxicology 5(2), 211-219.

Kratochwil NA, Huber W, Müller F, Kansy M & Gerber PR (2002). Predicting plasma protein binding of drugs: a new approach. Biochemical Pharmacology 64(9), 1355-1374.

Kriegl Jan M, Arnhold T, Beck B & Fox T (2005). Prediction of Human Cytochrome P450 Inhibition Using Support Vector Machines. QSAR & Combinatorial Science 24(4), 491-502.

Kuhn B, Jacobsen W, Christians U, Benet LZ & Kollman PA (2001). Metabolism of Sirolimus and Its Derivative Everolimus by Cytochrome P450 3A4: Insights from Docking, Molecular Dynamics, and Quantum Chemical Calculations. Journal of Medicinal Chemistry 44(12), 2027-2034.

Lanevskij K, Japertas P, Didziapetris R & Petrauskas A (2009). Ionization-specific prediction of blood-brain permeability. Journal of Pharmaceutical Sciences 98(1), 122-134.

Langowski J & Long A (2002). Computer systems for the prediction of xenobiotic metabolism. Advanced Drug Delivery Reviews 54(3), 407-415.

Lewis DF, Ioannides C & Parke DV (1996). COMPACT and molecular structure in toxicity assessment: a prospective evaluation of 30 chemicals currently being tested for rodent carcinogenicity by the NCI/NTP. Environmental Health Perspectives 104, 1011–1016.

Lewis DFV (2001). COMPACT: a structural approach to the modelling of cytochromes P450 and their interactions with xenobiotics. Journal of Chemical Technology & Biotechnology 76(3), 237-244.

Lewis DFV (2002). Molecular modeling of human cytochrome P450–substrate interactions. Drug Metabolism Reviews 34(1), 55 - 67.

Lewis DFV, Bird MG, Dickins M, Lake BG, Eddershaw PJ, Tarbit MH & Goldfarb PS (2000). Molecular modelling of human CYP2E1 by homology with the CYP102 haemoprotein domain: investigation of the interactions of substrates and inhibitors within the putative active site of the human CYP2E1 isoform. Xenobiotica 30(1), 1-25.

Lewis DFV, Eddershaw PJ, Goldfarb PS & Tarbit MH (1996a). Molecular modelling of CYP3A4 from an alignment with CYP102: Identification of key interactions between putative active site residues and CYP3A-specific chemicals. Xenobiotica 26(10), 1067-1086.

Lewis DFV, Eddershaw PJ, Goldfarb PS & Tarbit MH (1997). Molecular modelling of cytochrome P4502D6 (CYP2D6) based on an alignment with CYP102: structural studies on specific CYP2D6 substrate metabolism. Xenobiotica 27(4), 319-340.

Lewis DFV & Lake BG (1996). Molecular modelling of CYP1A subfamily members based on an alignment with CYP102: rationalization of CYP1A substrate specificity in terms of active site amino acid residues. Xenobiotica 26(7), 723-753.

Lewis DFV & Lake BG (1998). Molecular modelling and quantitative structure-activity relationship studies on the interaction of omeprazole with cytochrome P450 isozymes. Toxicology 125(1), 31-44.

Lewis DFV, Lake BG, Dickins M & Goldfarb PS (2004). Homology modelling of CYP3A4 from the CYP2C5 crystallographic template: analysis of typical CYP3A4 substrate interactions. Xenobiotica 34(6), 549-569.

Lewis DFV, Modi S & Dickins M (2002). Structure–activity relationship for human cytochrome P450 substrates and inhibitors. Drug Metabolism Reviews 34(1), 69 - 82.

Li H, Sun J, Fan X, Sui X, Zhang L, Wang Y & He Z (2008). Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction. Journal of Computer-Aided Molecular Design 22(11), 843-855.

Li H, Yap CW, Ung CY, Xue Y, Cao ZW & Chen YZ (2005). Effect of Selection of Molecular Descriptors on the Prediction of Blood• Brain Barrier Penetrating and Nonpenetrating Agents by Statistical Learning Methods. Journal of Chemical Information and Modeling 45(5), 1376-1384.

Lipinski C, Lombardo F, Dominy B & Feeney P (2001). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews 46, 3-26.

Lipinski CA, Lombardo F, Dominy BW & Feeney PJ (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews 23(1-3), 3-25.

Liu HX, Hu RJ, Zhang RS, Yao XJ, Liu MC, Hu ZD & Fan BT (2005). The prediction of human oral absorption for diffusion rate-limited drugs based on heuristic method and support vector machine. Journal of Computer-Aided Molecular Design 19(1), 33-46.

Liu R, Sun H & So S-S (2001). Development of Quantitative Structure• Property Relationship Models for Early ADME Evaluation in Drug Discovery. 2. Blood-Brain Barrier Penetration. Journal of Chemical Information and Computer Sciences 41(6), 1623-1632.

Livingstone DJ & van de Waterbeemnd H (2008). In silico prediction of human bioavailability, in Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability (Methods and Principles in Medicinal Chemistry 40). H van de Waterbeemd, B Testa, R Mannhold, H Kubinyi, G Folkers (Eds). pp. 433-451.

Lobell M, Hendrix M, Hinzen B, Keldenich J, Meier H, Schmeck C, Schohe-Loop R, Wunberg T & Hillisch A (2006). In Silico ADMET Traffic Lights as a Tool for the Prioritization of HTS Hits. ChemMedChem 1(11), 1229-1236.

Lobell M, Molnár L & Keserü GM (2003). Recent advances in the prediction of blood-brain partitioning from molecular structure. Journal of Pharmaceutical Sciences 92(2), 360-370.

Lobell M & Sivarajah V (2003). In silico prediction of aqueous solubility, human plasma protein binding and volume of distribution of compounds from calculated pKa and AlogP98 values. Molecular Diversity 7(1), 69-87.

Locuson CW, Rock DA & Jones JP (2004). Quantitative Binding Models for CYP2C9 Based on Benzbromarone Analogues. Biochemistry 43(22), 6948-6958.

Lombardo F, Blake JF & Curatolo WJ (1996). Computation of Brain• Blood Partitioning of Organic Solutes via Free Energy Calculations. Journal of Medicinal Chemistry 39(24), 4750-4755.

Long AW, JD (2003). Quantitative structure–activity relationships for predicting metabolism and modeling cytochrome P450 enzyme activities. Environmental Toxicology and Chemistry 22(8), 1894-1899.

Luan F, Liu H, Gao Y, Guo L, Zhang X & Guo Y (2009). A Quantitative Structure-Activity Relationship Study of Some Commercially Available Cephalosporins. QSAR & Combinatorial Science 28(9), 1003-1009.

Luco JM (1999). Prediction of the Brain• Blood Distribution of a Large Set of Drugs from Structurally Derived Descriptors Using Partial Least-Squares (PLS) Modeling. Journal of Chemical Information and Computer Sciences 39(2), 396-404.

Luco JM & Marchevsky E (2006). QSAR Studies on Blood-Brain Barrier Permeation. Current Computer-Aided Drug Design 2, 31-55.

Ma C-Y, Yang S-Y, Zhang H, Xiang M-L, Huang Q & Wei Y-Q (2008). Prediction models of human plasma protein binding rate and oral bioavailability derived by using GA-CG-SVM method. Journal of Pharmaceutical and Biomedical Analysis 47(4-5), 677-682.

Madden JC (2010). In Silico Approaches For Predicting Adme Properties, Chapter 10 in *Recent Advances in QSAR Studies: Methods and Applications.* T Puzyn, J Leszczynski & MTD Cronin (Eds). Springer, Heidelberg, Germany, pp. 283-301.

Ma X-l, Chen C & Yang J (2005). Predictive model of blood-brain barrier penetration of organic compounds. Acta Pharmacologica Sinica 26(4), 500-512.

Manga N, Duffy JC, Rowe, PH & Cronin, MT (2005). Structure-based methods for the prediction of the dominant P450 enzyme in human drug biotransformation: consideration of CYP3A4, CYP2C9, CYP2D6. SAR and QSAR in Environmental Research 16(1-2), 43-61.

Manga N, Duffy JC, Rowe PH & Cronin MTD (2003). A hierarchical QSAR model for urinary excretion of drugs in humans as a predictive tool for biotransformation QSAR & Combinatorial Science 22(2), 263-273.

Martin YC (2005). A Bioavailability Score. Journal of Medicinal Chemistry 48(9), 3164-3170.

Merkwirth C, Mauser H, Schulz-Gasch T, Roche O, Stahl M & Lengauer T (2004). Ensemble Methods for Classification in Cheminformatics. Journal of Chemical Information and Computer Sciences 44(6), 1971-1978.

Meskin MS & Lien EJ (1985). QSAR analysis of drug excretion into human breast milk. Journal of Clinical and Hospital Pharmacy 10(3), 269-278.

Moda TL, Montanari CA & Andricopulo AD (2007a). Hologram QSAR model for the prediction of human oral bioavailability. Bioorganic & Medicinal Chemistry 15(24), 7738-7745.

Moda TL, Montanari CA & Andricopulo AD (2007b). In Silico Prediction of Human Plasma Protein Binding Using Hologram QSAR. Letters in Drug Design & Discovery 4, 502-509.

Moda TL, Torres LG, Carrara AE & Andricopulo AD (2008). PK/DB: database for pharmacokinetic properties and predictive in silico ADME models. Bioinformatics 24(19), 2270-2271.

Modi S, Paine MJ, Sutcliffe MJ, Lian LY, Primrose WU, Wolf CR & Roberts GCK (1996). A Model for Human Cytochrome P450 2D6 Based on Homology Modeling and NMR Studies of Substrate Binding. Biochemistry 35(14), 4540-4550.

Mohan C, Ghandi T, Garg D & Shinde R (2007). Computer-assisted methods in chemical toxicity prediction. Mini Reviews in Medicinal Chemistry 7(5), 499-507.

Molnár L & Keseru GM (2002). A neural network based virtual screening of cytochrome P450 3A4 inhibitors. Bioorganic & Medicinal Chemistry Letters 12(3), 419-421.

Moon T, Chi MH, Kim D-H, Yoon CN & Choi Y-S (2000). Quantitative Structure-Activity Relationships (QSAR) Study of Flavonoid Derivatives for Inhibition of Cytochrome P450 1A2. Quantitative Structure-Activity Relationships 19(3), 257-263.

Narayanan R & Gunturi SB (2005). In silico ADME modelling: prediction models for blood-brain barrier permeation using a systematic variable selection method. Bioorganic & Medicinal Chemistry 13(8), 3017-3028.

Ng C, Xiao Y, Putnam W, Lum B & Tropsha A (2004). Quantitative structure-pharmacokinetic parameters relationships (QSPKR) analysis of antimicrobial agents in humans using simulated annealing k-nearest-neighbor and partial least-square analysis methods. Journal of Pharmaceutical Sciences 93(10), 2535-2544.

Niwa T (2003). Using General Regression and Probabilistic Neural Networks To Predict Human Intestinal Absorption with Topological Descriptors Derived from Two-Dimensional Chemical Structures. Journal of Chemical Information and Computer Sciences 43(1), 113-119.

Norinder U & Bergström Christel AS (2006). Prediction of ADMET Properties. ChemMedChem 1(9), 920-937.

Norinder U & Haeberlein M (2002). Computational approaches to the prediction of the blood-brain distribution. Advanced Drug Delivery Reviews 54(3), 291-313.

Norinder U & Österberg T (2001). Theoretical calculation and prediction of drug transport processes using simple parameters and partial least squares projections to latent structures (PLS) statistics. The use of electrotopological state indices. Journal of Pharmaceutical Sciences 90(8), 1076-1085.

Norinder U, Österberg T & Artursson P (1999). Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics. European Journal of Pharmaceutical Sciences 8(1), 49-56.

Norinder U, Sjöberg P & Österberg T (1998). Theoretical calculation and prediction of brain-blood partitioning of organic solutes using MolSurf parametrization and PLS statistics. Journal of Pharmaceutical Sciences 87(8), 952-959.

O'Brien SE & de Groot MJ (2005). Greater Than the Sum of Its Parts: Combining Models for Useful ADMET Prediction. Journal of Medicinal Chemistry 48(4), 1287-1291.

Obrezanova O, Csanyi G, Gola JMR & Segall MD (2007). Gaussian Processes: A Method for Automatic QSAR Modeling of ADME Properties. Journal of Chemical Information and Modeling 47(5), 1847-1857.

Obrezanova O, Gola J, Champness E & Segall M (2008). Automatic QSAR modeling of ADME properties: blood–brain barrier penetration and aqueous solubility. Journal of Computer-Aided Molecular Design 22(6), 431-440.

Ooms F, Weber P, Carrupt P-A & Testa B (2002). A simple model to predict blood-brain barrier permeation from 3D molecular fields. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1587(2-3), 118-125.

Oprea TI & Gottfries J (1999). Toward minimalistic modeling of oral drug absorption. Journal of Molecular Graphics and Modelling 17(5-6), 261-274.

Osterberg T & Norinder U (2000). Prediction of Polar Surface Area and Drug Transport Processes Using Simple Parameters and PLS Statistics. Journal of Chemical Information and Computer Sciences 40(6), 1408-1411.

Otagiri (2005). A molecular functional study on the interactions of drugs with plasma proteins. Drug Metabolism and Pharmacokinetics 20(5), 309-323.

Paixäo P, Gouveia LF & Morais JAG (2010). Prediction of the in vitro intrinsic clearance determined in suspensions of human hepatocytes by using artificial neural networks. European Journal of Pharmaceutical Sciences 39(5), 310-321.

Palm K, Stenberg P, Luthman K & Artursson P (1997). Polar Molecular Surface Properties Predict the Intestinal Absorption of Drugs in Humans. Pharmaceutical Research 14(5), 568-571.

Pan D, Iyer M, Liu J, Li Y & Hopfinger AJ (2004). Constructing Optimum Blood Brain Barrier QSAR Models Using a Combination of 4D-Molecular Similarity Measures and Cluster Analysis. Journal of Chemical Information and Computer Sciences 44(6), 2083-2098.

Pérez MAC, Sanz MB, Torres LR, Ávalos RG, González MP & Díaz HG (2004). A topological sub-structural approach for predicting human intestinal absorption of drugs. European Journal of Medicinal Chemistry 39(11), 905-916.

Pintore M, van de Waterbeemd H, Piclin N & Chrétien JR (2003). Prediction of oral bioavailability by adaptive fuzzy partitioning. European Journal of Medicinal Chemistry 38(4), 427-431.

Platts JA, Abraham MH, Zhao YH, Hersey A, Ijaz L & Butina D (2001). Correlation and prediction of a large blood-brain distribution data set--an LFER study. European Journal of Medicinal Chemistry 36(9), 719-730.

Polley MJ, Burden FR & Winkler DA (2005). Predictive Human Intestinal Absorption QSAR Models Using Bayesian Regularized Neural Networks. Australian Journal of Chemistry 58(12), 859-863.

Poso A, Gynther J & Juvonen R (2001). A comparative molecular field analysis of cytochrome P450 2A5 and 2A6 inhibitors. Journal of Computer-Aided Molecular Design 15(3), 195-202.

Raevsky OA, Fetisov VI, Trepalina EP, McFarland JW & Schaper K-J (2000). Quantitative Estimation of Drug Absorption in Humans for Passively Transported Compounds on the Basis of Their Physico-chemical Parameters. Quantitative Structure-Activity Relationships 19(4), 366-374.

Raevsky OA, Schaper K-J, Artursson P & McFarland JW (2002). A Novel Approach for Prediction of Intestinal Absorption of Drugs in Humans based on Hydrogen Bond Descriptors and Structural Similarity. Quantitative Structure-Activity Relationships 20(5-6), 402-413.

Rao S, Aoyama R, Schrag M, Trager WF, Rettie A & Jones JP (2000). A Refined 3-Dimensional QSAR of Cytochrome P450 2C9: Computational Predictions of Drug Interactions. Journal of Medicinal Chemistry 43(15), 2789-2796.

Rodgers SL, Davis AM, Tomkinson NP & van de Waterbeemd H (2007b). QSAR Modeling Using Automatically Updating Correction Libraries: Application to a Human Plasma Protein Binding Model. Journal of Chemical Information and Modeling 47(6), 2401-2407.

Rodgers SLD, Andrew M.; van de Waterbeemd, Han (2007a). Time-series QSAR analysis of human plasma protein binding data QSAR & Combinatorial Science 26(4), 511-521.

Rose K, Hall LH & Kier LB (2002). Modeling Blood-Brain Barrier Partitioning Using the Electrotopological State. Journal of Chemical Information and Computer Sciences 42(3), 651-666.

Saiakhov R, Stefan L & Klopman G (2000). Multiple computer-automated structure evaluation model of the plasma protein binding affinity of diverse drugs. Perspectives in Drug Discovery and Design 19(1), 133-155.

Salminen T, Pulli A & Taskinen J (1997). Relationship between immobilised artificial membrane chromatographic retention and the brain penetration of structurally diverse drugs. Journal of Pharmaceutical and Biomedical Analysis 15(4), 469-477.

Schuster D, Steindl TM & Langer T (2006). Predicting Drug Metabolism Induction In Silico. Current Topics in Medicinal Chemistry 6, 1627-1640.

Segall MD, Beresford AP, Gola JM, Hawksley D & Tarbit MH (2006). Focus on success: using a probabilistic approach to achieve an optimal balance of compound properties in drug discovery. Expert Opinion on Drug Metabolism & Toxicology 2(2), 325-337.

Shen M, Xiao Y, Golbraikh A, Gombar VK & Tropsha A (2003). Development and Validation of k-Nearest-Neighbor QSPR Models of Metabolic Stability of Drug Candidates. Journal of Medicinal Chemistry 46(14), 3013-3020.

Sheridan RP, Korzekwa KR, Torres RA & Walker MJ (2007). Empirical Regioselectivity Models for Human Cytochromes P450 3A4, 2D6, and 2C9. Journal of Medicinal Chemistry 50(14), 3173-3184.

Sietsema WK (1989). The absolute oral bioavailability of selected drugs International Journal of Clinical Pharmacology, Therapy and Toxicology 27(4), 179-211.

Singh SB, Shen LQ, Walker MJ & Sheridan RP (2003). A Model for Predicting Likely Sites of CYP3A4-mediated Metabolism on Drug-like Molecules. Journal of Medicinal Chemistry 46(8), 1330-1336.

Smith PA, Sorich MJ, McKinnon RA & Miners JO (2003). Pharmacophore and Quantitative Structure• Activity Relationship Modeling: Complementary Approaches for the Rationalization and Prediction of UDP-Glucuronosyltransferase 1A4 Substrate Selectivity. Journal of Medicinal Chemistry 46(9), 1617-1626.

Snyder R, Sangar R, Wang J & Ekins S (2002). Three-Dimensional Quantitative Structure Activity Relationship for Cyp2d6 Substrates. Quantitative Structure-Activity Relationships 21(4), 357-368.

Soffers AEMF, Ploemen JHTM, Moonen MJH, Wobbes T, van Ommen B, Vervoort J, van Bladeren PJ & Rietjens IMCM (1996). Regioselectivity and Quantitative Structure• Activity Relationships for the Conjugation of a Series of Fluoronitrobenzenes by Purified Glutathione S-Transferase Enzymes from Rat and Man. Chemical Research in Toxicology 9(3), 638-646.

Sorich MJ, McKinnon RA, Miners JO, Winkler DA & Smith PA (2004). Rapid Prediction of Chemical Metabolism by Human UDP-glucuronosyltransferase Isoforms Using Quantum Chemical Descriptors Derived with the Electronegativity Equalization Method. Journal of Medicinal Chemistry 47(21), 5311-5317.

Sorich MJ, Miners JO, McKinnon RA, Winkler DA, Burden FR & Smith PA (2003). Comparison of Linear and Nonlinear Classification Algorithms for the Prediction of Drug and Chemical Metabolism by Human UDP-Glucuronosyltransferase Isoforms. Journal of Chemical Information and Computer Sciences 43(6), 2019-2024.

Sorich MJ, Smith PA, McKinnon RA & Miners JO (2002). Pharmacophore and quantitative structure activity relationship modelling of UDP-glucuronosyltransferase 1A1 (UGT1A1) substrates. Pharmacogenetics and Genomics 12(8), 635-645.

Sprous DG, Palmer RK, Swanson JT & Lawless M (2010). QSAR in the Pharmaceutical Research Setting: QSAR Models for Broad, Large Problems. Current Topics in Medicinal Chemistry 10, 619-637.

Stanton DT, Mattioni BE, Knittel JJ & Jurs PC (2004). Development and Use of Hydrophobic Surface Area (HSA) Descriptors for Computer-Assisted Quantitative Structure• Activity and Structure• Property Relationship Studies. Journal of Chemical Information and Computer Sciences 44(3), 1010-1023.

Stoner CL, Cleton A, Johnson K, Oh D-M, Hallak H, Brodfuehrer J, Surendran N & Han H-K (2004). Integrated oral bioavailability projection using in vitro screening data as a selection tool in drug discovery. International Journal of Pharmaceutics 269(1), 241-249.

Strobl GR, von Kruedener S, Stoeckigt J, Guengerich FP & Wolff T (1993). Development of a pharmacophore for inhibition of human liver cytochrome P-450 2D6: molecular modeling and inhibition studies. Journal of Medicinal Chemistry 36(9), 1136-1145.

Subramanian G & Kitchen D (2003). Computational models to predict blood–brain barrier permeation and CNS activity. Journal of Computer-Aided Molecular Design 17(10), 643-664.

Subramanian G & Kitchen D (2006). Computational approaches for modeling human intestinal absorption and permeability. Journal of Molecular Modeling 12(5), 577-589.

Subramanian K (2005). TruPK – human pharmacokinetic models for quantitative ADME prediction. Expert Opinion on Drug Metabolism & Toxicology 1(3), 555-564.

Sugawara M, Takekuma Y, Yamada H, Kobayashi M, Iseki K & Miyazaki K (1998). A general approach for the prediction of the intestinal absorption of drugs: Regression analysis using the physicochemical properties and drug-membrane electrostatic interaction. Journal of Pharmaceutical Sciences 87(8), 960-966.

Sun H (2004). A Universal Molecular Descriptor System for Prediction of LogP, LogS, LogBB, and Absorption. Journal of Chemical Information and Computer Sciences 44(2), 748-757.

Sun LZ, Ji ZL, Chen X, Wang JF & Chen YZ (2002). ADME-AP: a database of ADME associated proteins. Bioinformatics 18(12), 1699-1700.

Sushko I, Novotarskyi S, Körner R, Pandey AK, Kovalishyn VV, Prokopenko VV & Tetko IV (2010). Applicability domain for in silico models to achieve accuracy of experimental measurements. Journal of Chemometrics 24(3-4), 202-208.

Susnow RG & Dixon SL (2003). Use of Robust Classification Techniques for the Prediction of Human Cytochrome P450 2D6 Inhibition. Journal of Chemical Information and Computer Sciences 43(4), 1308-1315.

Talafous J, Sayre L, Mieyal J & Klopman G (1994). META. 2. A dictionary model of mammalian xenobiotic metabolism. Journal of Chemical Information and Computer Sciences 34, 1326-1333.

Terfloth L, Bienfait B & Gasteiger J (2007). Ligand-Based Models for the Isoform Specificity of Cytochrome P450 3A4, 2D6, and 2C9 Substrates. Journal of Chemical Information and Modeling 47(4), 1688-1701.

Testa B, Balmat A-L, Long A & Judson P (2005a). Predicting Drug Metabolism - An Evaluation of the Expert System METEOR. Chemistry & Biodiversity 2(7), 872-885.

Testa B, Vistoli G & Pedretti A (2005b). Musings on ADME Predictions and Structure-Activity Relations. Chemistry & Biodiversity 2(11), 1411-1427.

Tetko IV, Bruneau P, Mewes HW, Rohrer DC & Poda GI (2006). Can we estimate the accuracy of ADME-Tox predictions? Drug Discovery Today 11(15-16), 700-707.

Thummel KE, Shen GG (2001). Design and optimization of dosage regimens: Pharmacokinetic data. Hardman J, Limbird E (eds); Goodman & Gilman's The Pharmacological Basis of Therapeutics; McGraw-Hill, New York, 1917–2023.

Thummel KE & Shen DD (2001). Design and optimization of dosage regimens: pharmacokinetic data. In: Goodman and Gilman's the pharmacological basis of therapeutics (Hardman JG, Limbird LE, Gilman AG, eds). McGraw-Hill, New York, pp. 1917-2023.

Tintori C, Magnani M, Schenone S & Botta M (2009). Docking, 3D-QSAR studies and in silico ADME prediction on c-Src tyrosine kinase inhibitors. European Journal of Medicinal Chemistry 44(3), 990-1000.

Trainor GL (2007). The importance of plasma protein binding in drug discovery Expert Opinion on Drug Discovery 2(1), 51-64.

Turner JV, Glass BD & Agatonovic-Kustrin S (2003a). Prediction of drug bioavailability based on molecular structure. Analytica Chimica Acta 485(1), 89-102.

Turner J, Maddalena D & Agatonovic-Kustrin S (2004a). Bioavailability Prediction Based on Molecular Structure for a Diverse Series of Drugs. Pharmaceutical Research 21(1), 68-82.

Turner JV, Maddalena DJ & Cutler DJ (2004b). Pharmacokinetic parameter prediction from drug structure using artificial neural networks. International Journal of Pharmaceutics 270(1-2), 209-219.

Turner JV, Maddalena DJ, Cutler DJ & Agatonovic-Kustrin S (2003b). Multiple pharmacokinetic parameter prediction for a series of cephalosporins. Journal of Pharmaceutical Sciences 92(3), 552-559.

Valko K, Nunhuck S, Bevan C, Abraham MH & Reynolds DP (2003). Fast gradient HPLC method to determine compounds binding to human serum albumin. Relationships with octanol/water and immobilized artificial membrane lipophilicity. Journal of Pharmaceutical Sciences 92(11), 2236-2248.

Van Damme S, Langenaeker W & Bultinck P (2008). Prediction of blood-brain partitioning: A model based on ab initio calculated quantum chemical descriptors. Journal of Molecular Graphics and Modelling 26(8), 1223-1236.

van de Waterbeemd H & Kansy M (1992). Hydrogen-Bonding Capacity and Brain Penetration. CHIMIA International Journal for Chemistry 46, 299-303.

Vasanthanathan P, Olsen L, Jorgensen FS, Vermeulen NPE & Oostenbrink Chris (2010). Computational Prediction of Binding Affinity for CYP1A2-Ligand Complexes Using Empirical Free Energy Calculations. Drug Metabolism and Disposition (*Article in press*).

Vastag M & Keseru GM (2009). Current in vitro and in silico models of blood-brain barrier penetration: a practical view. Current Opinion in Drug Discovery & Development 12(1), 115-124.

Vaz RJ, Nayeem A, Santone K, Chandrasena G & Gavai AV (2005). A 3D-QSAR model for CYP2D6 inhibition in the aryloxypropanolamine series. Bioorganic & Medicinal Chemistry Letters 15(17), 3816-3820.

Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW & Kopple KD (2002). Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. Journal of Medicinal Chemistry 45(12), 2615-2623.

Verma R, Hansch C & Selassie C (2007). Comparative QSAR studies on PAMPA/modified PAMPA for high throughput profiling of drug absorption potential with respect to Caco-2 cells and human intestinal absorption. Journal of Computer-Aided Molecular Design 21(1), 3-22.

Veselovsky A, Sobolev B, Zharkova M & Archakov A (2010). Computer-based substrate specificity prediction for cytochrome P450. Biochemistry (Moscow) Supplemental Series B: Biomedical Chemistry 4(1), 75-81.

Vilar S, Chakrabarti M & Costanzi S (2010). Prediction of passive blood-brain partitioning: Straightforward and effective classification models based on in silico derived physicochemical descriptors. Journal of Molecular Graphics and Modelling (*Article in press*).

Votano J (2005). Recent uses of topological indices in the development of in silico ADMET models.Votano JR. Current Opinion in Drug Discovery & Development 8(1), 32-37.

Votano JR, Parham M, Hall LM, Hall LH, Kier LB, Oloff S & Tropsha A (2006). QSAR Modeling of Human Serum Protein Binding with Several Modeling Techniques Utilizing Structure• Information Representation. Journal of Medicinal Chemistry 49(24), 7169-7181.

Wan H & Ulander J (2006). High-throughput pKa screening and prediction amenable for ADME profiling. Expert Opinion on Drug Metabolism & Toxicology 2(1), 139-155.

Wanchana S, Yamashita F & Hashida M (2003). QSAR Analysis of the Inhibition of Recombinant CYP 3A4 Activity by Structurally Diverse Compounds Using a Genetic Algorithm-Combined Partial Least Squares Method. Pharmaceutical Research 20(9), 1401-1408.

Wang J & Skolnik S (2010). Mitigating permeability-mediated risks in drug discovery. Expert Opinion on Drug Metabolism & Toxicology 6(2), 171-187.

Wang J, Krudy G, Xie X-Q, Wu C & Holland G (2006). Genetic Algorithm-Optimized QSPR Models for Bioavailability, Protein Binding, and Urinary Excretion. Journal of Chemical Information and Modeling 46(6), 2674-2683.

Wang Q & Halpert JR (2002). Combined Three-Dimensional Quantitative Structure-Activity Relationship Analysis of Cytochrome P450 2B6 Substrates and Protein Homology Modeling. Drug Metabolism and Disposition 30(1), 86-95.

Weaver S & Gleeson MP (2008). The importance of the domain of applicability in QSAR modeling. Journal of Molecular Graphics and Modelling 26(8), 1315-1326.

Wegner JK, Frohlich H & Zell A (2004). Feature Selection for Descriptor Based Classification Models. 2. Human Intestinal Absorption (HIA). Journal of Chemical Information and Computer Sciences 44(3), 931-939.

Wessel MD, Jurs PC, Tolan JW & Muskal SM (1998). Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure. Journal of Chemical Information and Computer Sciences 38(4), 726-735.

Wichmann K, Diedenhofen M & Klamt A (2007). Prediction of Blood-• rain Partitioning and Human Serum Albumin Binding Based on COSMO-RS • -Moments. Journal of Chemical Information and Modeling 47(1), 228-233.

Winkler DA & Burden FR (2004). Modelling blood-brain barrier partitioning using Bayesian neural nets. Journal of Molecular Graphics and Modelling 22(6), 499-505.

Wolohan PRN & Clark RD (2003). Predicting drug pharmacokinetic properties using molecular interaction fields and SIMCA. Journal of Computer-Aided Molecular Design 17(1), 65-76.

Xue CX, Zhang RS, Liu HX, Yao XJ, Liu MC, Hu ZD & Fan BT (2004a). QSAR Models for the Prediction of Binding Affinities to Human Serum Albumin Using the Heuristic Method and a Support Vector Machine. Journal of Chemical Information and Computer Sciences 44(5), 1693-1700.

Xue Y, Li ZR, Yap CW, Sun LZ, Chen X & Chen YZ (2004b). Effect of Molecular Descriptor Feature Selection in Support Vector Machine Classification of Pharmacokinetic and Toxicological Properties of Chemical Agents. Journal of Chemical Information and Computer Sciences 44(5), 1630-1638.

Yamazaki K & Kanaoka M (2004). Computational prediction of the plasma protein-binding percent of diverse pharmaceutical compounds. Journal of Pharmaceutical Sciences 93(6), 1480-1494.

Yap CW & Chen YZ (2005a). Quantitative structure-pharmacokinetic relationships for drug distribution properties by using general regression neural network. Journal of Pharmaceutical Sciences 94(1), 153-168.

Yap CW & Chen YZ (2005b). Prediction of Cytochrome P450 3A4, 2D6, and 2C9 Inhibitors and Substrates by Using Support Vector Machines. Journal of Chemical Information and Modeling 45(4), 982-992.

Yap CW, Li ZR & Chen YZ (2006). Quantitative structure-pharmacokinetic relationships for drug clearance by using statistical learning methods. Journal of Molecular Graphics and Modelling 24(5), 383-395.

Yoshida F & Topliss JG (2000). QSAR Model for Drug Human Oral Bioavailability1. Journal of Medicinal Chemistry 43(13), 2575-2585.

Young RC, Mitchell RC, Brown TH, Ganellin CR, Griffiths R, Jones M, Rana KK, Saunders D & Smith IR (1988). Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H2 receptor histamine antagonists. Journal of Medicinal Chemistry 31(3), 656-671.

Zamora I, Afzelius L & Cruciani G (2003). Predicting Drug Metabolism: A Site of Metabolism Prediction Tool Applied to the Cytochrome P450 2C9. Journal of Medicinal Chemistry 46(12), 2313-2324.

Zhang H (2004). A new nonlinear equation for the tissue/blood partition coefficients of neutral compounds. Journal of Pharmaceutical Sciences 93(6), 1595-1604.

Zhang H (2005). A New Approach for the Tissue• Blood Partition Coefficients of Neutral and Ionized Compounds. Journal of Chemical Information and Modeling 45(1), 121-127.

Zhang H (2006). A QSAR Study of the Brain/Blood Partition Coefficients on the Basis of pKa Values. QSAR & Combinatorial Science 25(1), 15-24.

Zhang L, Zhu H, Oprea T, Golbraikh A & Tropsha A (2008a). QSAR Modeling of the Blood–Brain Barrier Permeability for Diverse Organic Compounds. Pharmaceutical Research 25(8), 1902-1914.

Zhang H, Hu S & Zhang Y (2008b). Prediction of Distribution of Neutral, Acidic and Basic Structurally Diverse Compounds Between Blood and Brain by the Nonlinear Methodology. Medicinal Chemistry 4, 170-189.

Zhang H & Zhang Y (2006). Convenient Nonlinear Model for Predicting the Tissue/Blood Partition Coefficients of Seven Human Tissues of Neutral, Acidic, and Basic Structurally Diverse Compounds. Journal of Medicinal Chemistry 49(19), 5815-5829.

Zhao YH, Abraham MH, Ibrahim A, Fish PV, Cole S, Lewis ML, de Groot MJ & Reynolds DP (2007). Predicting Penetration Across the Blood-Brain Barrier from Simple Descriptors and Fragmentation Schemes. Journal of Chemical Information and Modeling 47(1), 170-175.

Zhao YH, Abraham MH, Le J, Hersey A, Luscombe CN, Beck G, Sherborne B & Cooper I (2002). Rate-Limited Steps of Human Oral Absorption and QSAR Studies. Pharmaceutical Research 19(10), 1446-1457.

Zhao YH, Le J, Abraham MH, Hersey A, Eddershaw PJ, Luscombe CN, Boutina D, Beck G, Sherborne B, Cooper I & Platts JA (2001). Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. Journal of Pharmaceutical Sciences 90(6), 749-784.

Zhou D, Afzelius L, Grimm SW, Andersson TB, Zauhar RJ & Zamora I (2006). Comparison of methods for the prediction of the metabolic sites for CYP3A4-mediated metabolic reactions. Drug Metabolism and Disposition 34(6), 976-983.

Zhou D, Liu R, Otmani SA, Grimm SW, Zauhar RJ & Zamora I (2007). Rapid Classification of CYP3A4 Inhibition Potential Using Support Vector Machine Approach. Letters in Drug Design & Discovery 4, 192-200.

Zmuidinavicius D, Didziapetris R, Japertas P, Avdeef A & Petrauskas A (2003). Classification structure-activity relations (C-SAR) in prediction of human intestinal absorption. Journal of Pharmaceutical Sciences 92(3), 621-633.

Zuegge J, Fechner U, Roche O, Parrott Neil J, Engkvist O & Schneider G (2002). A fast virtual screening filter for cytochrome P450 3A4 inhibition liability of compound libraries. Quantitative Structure-Activity Relationships 21(3), 249-256.

TABLES

Table 1. Recent (2005-2010) reviews/expert opinions concerning *in silico* studies in ADME and ADME-related endpoints

Year	Reference
2010	Madden (2010). In silico approaches for predicting ADME properties
2010	Veselovsky et al. (2010). Computer-based substrate specificity prediction for cytochrome P450
2010	Wang & Skolnik (2010). Mitigating permeability-mediated risks in drug discovery
2010	Kortagere & Ekins (2010). Troubleshooting computational methods in drug discovery
2010	Cross & Cruciani (2010). Molecular fields in drug discovery: getting old or reaching maturity?
2010	Cruciani et al. (2010). ChemInform abstract: *In silico* pKa prediction and ADME profiling
2010	Sprous et al. (2010). QSAR in the pharmaceutical research setting: QSAR models for broad, large problems
2010	Kharkar (2010). Two-Dimensional (2D) *in silico* models for Absorption, Distribution, Metabolism, Excretion and Toxicity (ADME/T) in drug discovery
2010	Ekins (2010). Precompetitive preclinical ADME/Tox data: set it free on the web to facilitate computational model building and assist drug development
2009	Franklin (2009). *In silico* studies in ADME/Tox: caveat emptor
2009	Livingstone & van de Waterbeemd (2009). *In silico* prediction of human bioavailability
2009	Vastag & Keserü (2009). Current *in vitro* and *in silico* models of BBB penetration: a practical view
2008	Chohan et al. (2008). Advancements in predictive *in silico* models for ADME
2008	Hou & Wang (2008). Structure – ADME relationship: still a long way to go?
2008	Jacobs et al. (2008). The use of metabolising systems for *in vitro* testing of endocrine disruptors
2008	Li et al. (2008). Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction
2007	Clark (2007). *In silico* ADMET tools: a dawn of a new generation?
2007	Dearden & Worth (2007). *In silico* prediction of physicochemical properties
2007	Dearden (2007). *In silico* prediction of ADMET properties: how far have we come?
2007	Khan & Sylte (2007). Predictive QSAR modeling for the successful predictions of the ADMET properties of candidate drug molecules
2007	Mohan et al. (2007). Computer-assisted methods in chemical toxicity prediction
2007	Al-Fahemi et al. (2007). Investigating the utility of momentum-space descriptors for predicting BBB penetration
2007	Ekins et al. (2007a). *In silico* pharmacology for drug discovery: methods for virtual ligand screening and profiling
2007	Ekins et al. (2007b). *In silico* pharmacology for drug discovery: applications to targets and beyond
2007	Ekins et al. (2007c). Novel applications of kernel-PLS to modeling a comprehensive array of properties for drug discovery
2007	Trainor (2007). The importance of plasma protein binding in drug discovery
2006	Hou et al. (2006). Recent advances in computational prediction of drug bbsorption and permeability in drug discovery

Year	Reference
2006	Chohan et al. (2006). Quantitative Structure Activity Relationships in drug metabolism
2006	Crivori & Pogessi (2006). Computational approaches for predicting CYP-related metabolism properties in the screening of new drugs
2006	Fox & Kriegl (2006). Machine learning techniques for *in silico* modeling of drug metabolism
2006	Norinder & Bergström (2006). Prediction of ADMET properties
2006	Gola et al. (2006). ADMET property prediction: The state of the art and current challenges
2006	Schuster et al. (2006). Predicting drug metabolism induction *in silico*
2006	Tetko et al. (2006). Can we estimate the accuracy of ADME-Tox predictions?
2006	Wang & Ulander (2006). High-throughput pKa screening and prediction amenable for ADME profiling
2006	Segall et al. (2006). Focus on success: using a probabilistic approach to achieve an optimal balance of compound properties in drug discovery
2006	Hyland et al. (2006). Utility of human/human-derived reagents in drug discovery and development: An industrial perspective
2006	Luco & Marchevsky (2006). QSAR studies on blood-brain barrier permeation
2006	Allen & Geldenhuys (2006). Molecular modeling of blood–brain barrier nutrient transporters: *In silico* basis for evaluation of potential drug delivery to the central nervous system
2006	Cianchetta et al. (2006). Molecular Interaction Fields in ADME and safety
2005	Colmenarejo (2005). *In silico* ADME prediction: Data sets and models
2005	De Graaf et al. (2005). Cytochrome P450 *in silico*: an integrative modeling approach
2005	Delisle et al. (2005). Computational ADME/Tox modeling: aiding understanding and enhancing decision making in drug design
2005	Goodwin & Clark (2005). *In silico* predictions of BBB penetration: considerations to “keep in mind”
2005	Ekins et al. (2005). Computational prediction of human drug metabolism
2005	Ekins et al. (2005). Techniques: Application of systems biology to absorption, distribution, metabolism, excretion and toxicity
2005	Kaznessis (2005). A review of methods for computational prediction of BB partitioning
2005	Otagiri (2005). A molecular functional study on the interactions of drugs with plasma proteins
2005	Testa et al. (2005b). Musings on ADME predictions and structure-activity relations
2005	Votano (2005). Recent uses of topological indices in the development of *in silico* ADMET models

Table 2. Databases for ADME

Database (developer and availability)	Database size/ chemical classes	Provided properties	Details
ADME INDEX™ DATABASE Bio-Rad Laboratories http://www.bio-rad.com/ (commercial; hosted by Bio-Rad Lab KnowItAll)	FDA-approved drugs, non-approved compounds	ADME	Experimental *in vitro* ADME data generated by Lighthouse Data Solutions Labs (LDS)
ADME DB Fujitsu http://www.fqs.pl/ (commercial, available online)	Drugs	Drug metabolizing enzymes, kinetic metabolism, transporters	Protein information about enzymes and transporters, metabolic reactions, types of drug-drug interactions, structures of drugs and metabolites, kinetic information
ADME-associated proteins (ADME-AP) DB Bio Info & Drug Design (Sun et al., 2002) http://xin.cz3.nus.edu.sg/group/admeap/admeap.asp/ (freely available online)	321 proteins, 964 substrates	ADME	Drug ADME-associated proteins, functions, similarities, substrates/ ligands, and tissue distributions
AurSCOPE® ADME/DDI Aureus Pharma http://www.aureus-pharma.com/ (commercial)	7000 compounds	ADME, Drug-drug interactions	Biological and chemical information on metabolic properties of drugs
BioPath Database Molecular Networks http://www.molecular-networks.com/ (trial version freely available online, commercial full version)	Endogenous compounds, 1175 chemicals in free version, 2074 chemicals in commercial version	1545 biochemical transformations (in free version), 2881 biochemical transformations (in commercial version)	Biochemical pathways (metabolic transformations and cellular regulations) for prokaryotes, plants, yeasts and animals; subcellular localisation of pathways including: cytosol, chloroplasts, mitochondria, endoplasmatic reticulum, peroxysomes, endothelium of blood vessels, vascular muscle cell, animal extracellular matrix, nucleus, animal cell membrane, plant cell wall

Database (developer and availability)	Database size/ chemical classes	Provided properties	Details
BioPrint® CEREP http://www.cerep.fr/ (commercial)	2500 compounds	Pharmacology and ADME database	Chemical descriptors (structures, 2D and 3D); *in vitro* profiles; *in vivo* effects; enzyme/ solubility/ absorption assays
KEGG Database Kanehisa Laboratories, Kyoto University & University of Tokyo http://www.genome.jp/kegg/ (free for academic use; for other purposes available commercially under license agreement with Pathway Solutions Inc.) http://www.pathway.jp/licensing/commercial.html)	16 databases, 344 metabolic pathway maps, 9150 drugs, 1231 organisms, 16083 metabolites, 8064 biochemical reactions	Metabolism	KEGG metabolism information includes (among others) the following aspects: carbohydrate/ energy/ lipid/ nucleotide/ amino acid/ metabolism; biosynthesis of secondary metabolites; xenobiotic biodegradation and metabolism
Metabolism Database Accelrys http://accelrys.com/ (commercial)	69 241 records (drugs, agrochemicals, food additives, industrial & environmental chemicals)	Metabolism	Metabolism data for vertebrates, invertebrates and plants; data on pathways and related compounds
Metabolism & Transport Drug Interaction Database (DIDB) University of Washington http://www.druginteractioninfo.org/ (commercial)	Drugs	Pharmacokinetic data, Enzyme/transporter interactions	Drug interactions in humans; pharmacokinetic profiles of drugs
Metabolite™ Symyx http://www.symyx.com/ (commercial)	Xenobiotics and drugs	Metabolism	Metabolic paths and schemes; experimental data
PharmGKB Database Stanford University http://www.pharmgkb.org/ (freely available for research purposes)	Drugs, genes, pathways, diseases, information about people who have participated in pharmcogenomics research studies	Pharmacokinetic data	Clinical and basic pharmacokinetic and pharmacogenomic research in the cardiovascular, pulmonary, cancer, pathways, metabolic and transporter domains

Database (developer and availability)	Database size/ chemical classes	Provided properties	Details
PharmaPendium™ Database Elsevier https://www.pharmapendium.com/ (commercial)	Data from the FDA freedom of information documents and EMEA "EPAR" approval documents (structure/substructure searchable)	Pharmacokinetic data	Data on efficacy, indications and dosage, safety, pharmacokinetics, pharmacology and mode of action, preclinical and clinical toxicity (extracted from documents), adverse effects (extracted from documents), general product information
PK/DB Database Moda et al., 2008 http://www.pkdb.ifsc.usp.br/ (freely available online)	1203 compounds	Pharmacokinetic data	Human intestinal absorption, human oral bioavailability, plasma protein binding, blood/brain barrier penetration
Prous Ensemble ® Database Prous Science http://www.prous.com/ (commercial)	127 000 bioactive compounds, 275 000 references	Pharmacokinetic and metabolism data	Drug monographs containing information on the synthesis, pharmacological actions, pharmacokinetics and metabolism, toxicity, clinical studies, manufacturers and references
Symcyp http://www.simcyp.com/ (commercial)	47 drugs – experimental data from *in vitro* enzyme and cellular systems, physicochemical properties and dosage forms	ADME, pharmacokinetic profiles, drug-drug interactions	Population-based PBPK simulator for modelling ADME and drug-drug interactions in virtual patient populations.
WOMBAT-PK 2009 Sunset Molecular http://www.sunsetmolecular.com/ (commercial)	1230 drugs	Pharmacokinetic data	Percentage oral bioavailability, percentage plasma protein binding, qualitative blood/brain barrier permeability, phase 1 metabolizing enzymes

Table 3. Literature datasets for ADME

Dataset (reference)	Dataset size/ chemical classes	ADME and related properties provided	Information available
Hou et al. (2007b)	648 compounds	Human intestinal absorption	HIA
Hou et al. (2007a)	768 compounds	Oral bioavailability	Oral bioavailability
Moda et al. (2007a)	302 drugs		
Sietsema et al. (1989)	Dataset & 550 references		
Konovalov et al. (2007)	328 compounds	Blood/brain barrier penetration	LogBB
Zhao et al. (2007)	1593 compounds		Binary classification (BBB+/BBB-)
Abraham et al. (2006)	328 drugs and organic compounds		Blood/plasma/serum to rat brain distribution coefficients
Li et al. (2005)	415 compounds		Binary classification (BBB+/BBB-)
Hollósy et al. (2006)	179 drugs	Plasma protein binding	Percentage PPB, urinary excretion and other ADME data
Votano et al. (2006)	1008 compounds		Percentage human plasma protein binding
Turner et al. (2004b)	62 drugs		Human plasma protein binding; total and renal clearance
Thummel & Shen (2001)	320 drugs		Percentage PPB, urinary excretion and other ADME data
Kalgutkar et al. (2005)	464 references	Metabolic pathways	Structural alerts
Manga et al. (2005)	147 drugs	CYP metabolism	CYP isoforms predominantly responsible for their metabolism (CYP3A4/2D6/2C9)
Yap et al. (2006)	503 compounds	Clearance (CL_{tot})	Total clearance in humans

Table 4. Software tools for predicting physicochemical properties useful as input data for ADME modelling

SOFTWARE (COMPANY)	AVAILABILITY	PROPERTY			
		Ionization constant (pKa)	Octanol/water partition coefficient (logP)	Distribution coefficient (logD)	Aqueous solubility ($logSa_q$)
ACD/PhysChem Suite/Batch (ACD Labs) http://www.acdlabs.com/	Commercial	•	•	•	•
ASTER (U.S. EPA) http://www.epa.gov/med/Prods_Pubs/aster.htm/	Not publicly available	•	•		•
ChemOffice (CambridgeSoft) http://www.cambridgesoft.com/	Commercial		•		
ChemProp (Helmholtz Centre for Environmental Research, UFZ) http://www.ufz.de/	Commercial		•		•
ClogP (DAYLIGHT) http://www.daylight.com/	Commercial		•		
EPISUITE (U.S. EPA) http://www.epa.gov/oppt/exposure/pubs/episuite.htm/	Freely downloadable		•		•
JAGUAR (Schrödinger) http://www.schrodinger.com/	Commercial	•			
Molecular Modeling Pro (ChemSW) http://www.chemsw.com/molecularmodeling.htm/	Commercial		•		•
MoKa (Molecular Discovery) http://www.moldiscovery.com/	Commercial	•			
Pipeline Pilot (Accelrys Scitegic) http://accelrys.com/	Commercial	•	•		•
SPARC (U.S. EPA) http://ibmlc2.chem.uga.edu/sparc/	Free on-line application	•	•		•
TSAR (Accelrys) http://accelrys.com/	Commercial		•		
VCCLAB (Virtual Computational Chemistry Lab) http://www.vcclab.org/	Free on-line application	•	•		•

Table 5. Software tools for physicochemical-based and organism-based (*) ADME predictions

SOFTWARE (COMPANY)	AVAILABILITY	PROPERTY											
		Ionization constant (pKa)	Octanol/water partition coefficient (logP)	Distribution coefficient (logD)	Aqueous solubility ($logS_q$)	GI absorption	Caco-2	Oral bioavailability	BBB	PPB	Metabolism	Transporters	Others (e.g. excretion)
ACD/ADME Suite with AbSolv module (ACD Labs) http://www.acdlabs.com/	Commercial	•	•	•	•	•	•	•	•	•	•		•
Accord for Excel with ADME/Tox Add-on (Accelrys) http://accelrys.com/	Commercial	•	•	•	•	•			•	•	•		
ADME Batches[1] (Pharma Algorithms) – now included in ACD/ADME Suite	Commercial				•	•							
ADME Boxes[1] (Pharma Algorithms) – now included in ACD/ADME Suite	Commercial	•	•	•	•	•	•	•		•		•	
DISCOVERY STUDIO including Cerius2 (Accelrys) http://accelrys.com/	Commercial				•	•			•	•	•		•
ADMENSA[1] (Inpharmatica)	Commercial		•		•	•		•		•	•		
ADMET Predictor (Simulations Plus Inc.) http://www.simulations-plus.com/	Commercial	•	•	•	•	•			•	•			•
ADMETox/Pallas including MetabolExpert, MEXAlert, pKalc, PrologD, TPSA, RetroMEX, RuleOf5, PrologP, ToxAlert, Cytotoxicity (CompuDrug) http://www.compudrug.com/	Commercial	•	•	•	•						•		•

SOFTWARE (COMPANY)	AVAILABILITY	PROPERTY											
		Ionization constant (pKa)	Octanol/water partition coefficient (logP)	Distribution coefficient (logD)	Aqueous solubility ($logSa_q$)	GI absorption	Caco-2	Oral bioavailability	BBB	PPB	Metabolism	Transporters	Others (e.g. excretion)
ADMEWORKS including Predictor and ModelBuilder (Fujitsu) http://www.fqs.pl/	Commercial		•		•	•			•		•		•
BioFrontier/P450 (Fujitsu) http://www.fqs.pl/	Commercial										•		
ChemDBsoft with MOLPRO Package including SLIPPER (ChemDBsoft) http://www.chemdbsoft.com/	Commercial	•	•	•	•	•							
ChemSilico Predictors, i.e. CS LogWS/D/P, CS BBB/PB/HIA (ChemSilico) http://chemsilico.com/	Commercial		•	•	•	•			•	•			
Cloe® including Cloe PK, Cloe PredictHIA (Cyprotex)* http://www.cyprotex.com/	Commercial					•					•		•
COMPACT (Computer-Optimised Molecular Parametric Analysis of Chemical Toxicity) University of Surrey, Guildford, UK Lewis et al. (1996, 2001)	In-house software, neither commercial nor public										•		
GastroPlus (Simulations Plus Inc.)* http://www.simulations-plus.com/	Commercial				•	•		•			•		
iDEA ADME[1] (Lion Biosciences)	Commercial				•	•	•	•	•		•		
iDEA PKexpress[1] (Lion Biosciences)	Commercial					•					•		
Jchem with Calculator Plugins (ChemAxon) http://www.chemaxon.com/	Commercial	•	•	•									•

SOFTWARE (COMPANY)	AVAILABILITY	PROPERTY											
		Ionization constant (pKa)	Octanol/water partition coefficient (logP)	Distribution coefficient (logD)	Aqueous solubility ($logSa_q$)	GI absorption	Caco-2	Oral bioavailability	BBB	PPB	Metabolism	Transporters	Others (e.g. excretion)
KnowItAll ADME/Tox (Bio-Rad Laboratories) http://www.bio-rad.com/	Commercial	•	•	•	•	•		•	•	•			•
META/METAPC/ MCASE ADME Module (MultiCASE) Klopman et al. (1994, 1997, 1999), Talafous et al. (1994) http://www.multicase.com/	Commercial							•		•	•		•
MetaDrug™ (Genego) http://www.genego.com/	Commercial		•	•	•				•	•	•	•	•
MetaSite (Molecular Discovery) Cruciani et al., 2005 http://www.moldiscovery.com/	Commercial										•		
METEOR (Lhasa Ltd.) Testa et al., 2005a http://www.lhasalimited.org/	Commercial										•		
MolCode ToolBox (MolCode) http://www.molcode.com/	Commercial						•		•	•	•		
NorayMet ADME (Noray Bioinformatics) http://www.noraybio.com/	Commercial	•	•	•	•	•	•			•	•		•
OraSpotter[1] (ZyxBio)	Commercial				•	•						•	
PK SiM (Bayer Technology Services) http://www.systems-biology.com/	Commercial					•		•			•	•	

SOFTWARE (COMPANY)	AVAILABILITY	PROPERTY											
		Ionization constant (pKa)	Octanol/water partition coefficient (logP)	Distribution coefficient (logD)	Aqueous solubility ($logSa_q$)	GI absorption	Caco-2	Oral bioavailability	BBB	PPB	Metabolism	Transporters	Others (e.g. excretion)
ProPred (CAPEC) http://www.capec.kt.dtu.dk/	Commercial		•		•								
PreADME (Bioinformatics and Molecular Design Research Centre), PreADMET web-based application (BMDRC) http://www.bmdrc.org/	Commercial		•		•	•	•		•	•			•
q-ADME (Quantum Lead) http://www.q-lead.com/						•	•			•			
QikProp (Schrödinger) http://www.schrodinger.com/	Commercial		•		•		•		•	•			•
QMPRPlus[1] (Simulations Plus Inc.) http://www.simulations-plus.com/	Commercial		•	•	•	•			•				
StarDrop (BioFocus DPI) http://www.scientific-computing.com/	Commercial	•	•	•	•	•			•	•	•	•	•
Simcyp® (SimCYP)* http://www.simcyp.com/	Commercial					•					•		
OASIS-TIMES (Laboratory of Mathematical Chemistry, Bourgas University) http://www.oasis-lmc.org/	Commercial										•		
TruPK[1] (Strand Genomics), now a part of KnowItAll platform from Bio-Rad Labs	Commercial					•				•	•		
VolSurf/VolSurf+ (Molecular Discovery & Tripos) http://www.moldiscovery.com/	Commercial		•	•	•	•	•		•	•	•		

[1] Former software, not commercially available now, but often cited and still possibly in use

Table 6. Rules-of-thumb developed for ADME

Reference	ADME property	Rules-of-thumb details
Gleeson (2008)	Solubility Bioavailability PPB Brain/tissue binding CYP1A2/2C9/2C1/2D6/3A4 inhibition	The influence of molecular weight (MW), ionization state (pKa) and calculated octanol/water partition coefficient (ClogP) on various ADME properties was discussed, e.g: • **Solubility** increases as: MW decreases and ClogP decreases. In terms of pKa: zwitterionic molecules containing both an acidic and basic functional group are the most highly soluble, while neutral molecules are the least soluble; acidic molecules are more soluble than basic molecules; • **Bioavailability** increases as MW decreases; ClogP does not have a significant influence. In terms of pKa: bioavailabilities for neutral, basic and zwitterionic molecules are quite similar; • **Plasma protein binding** increases as: MW increases and ClogP increases. In terms of pKa, PPB follows the trend: acids > neutrals > zwitterions > bases; • **Brain/tissue binding** increases as MW increases and ClogP increases. In terms of pKa, no significant relationships have been observed
Lobell et al. (2006)	GI absorption	Good GI absorption is characteristic for reasonably soluble, not too liphophilic, large, polar or flexible compounds. The combined calculated values of physicochemical properties determining these factors, i.e. aqueous solubility ($logS_{aq}$), octanol/water partition coefficient (ClogP), molecular weight (MW), polar surface area (PSA) and the number of rotatable bonds (RotB) give a "**traffic light" (TL) scheme for absorption**, as follows: • **Green**: $logS_{aq}$ • 50; ClogP • 3; MW • 400; PSA • 120; RotB • 7; • **Yellow**: $logS_{aq}$: 10-50; ClogP: 3-5; MW: 400-500; PSA: 120-140 RotB: 8-10; • **Red**: $logS_{aq}$ < 10; ClogP > 5; MW > 500; PSA > 140; RotB • 11
Zmuidinavicius et al. (2003)	Human intestinal absorption	• Compounds with quaternary nitrogens or biphosphonate moieties are poorly absorbed; • Compounds with molecular weight < 255 have good absorption; • Compounds with molecular weight between 255 and 580, polar surface area < 154 $Å^2$ and one of two following conditions hold: logP > 0 or hydrogen bond acidity < 1.3 display good absorption; • Compounds with molecular weight > 580, polar surface area < 291 $Å^2$ and logP > 0 are well absorbed
Norinder & Haberlein (2002)	BBB penetration	• The molecule has a high chance of entering the brain if the number of nitrogen and oxygen atoms (N+O) atoms is • 5; • LogBB is positive if [logP-(N+O)] is positive

Reference	ADME property	Rules-of-thumb details
Veber et al. (2002)	Oral bioavailability	High probability of good oral bioavailability for compounds with: • • 10 rotatable bonds; • Polar surface area • 140 A^2 or • The sum of hydrogen bond donors and acceptors • 12
Kelder et al. (1999)	BBB penetration	• The upper limit for the polar surface area (PSA) for a molecule that has a high chance of entering the brain is < 60-70 Å
Van der Waterbeemd et al. (1998)	BBB penetration	• The upper limit for the polar surface area (PSA) for a molecule that has a high chance of entering the brain is around 90 Å • The molecular weight (MW) of such molecule should be not larger than 450 g/mol
Lipinski et al. (1997)	Absorption	"Rule of 5", indicating that a molecule is prone to poor absorption if: • Molecular weight > 500; • Sum of OH and NH hydrogen bond donors > 5; • Sum of O and N hydrogen bond donors > 10; • ClogP > 5

Table 7. Literature models for human intestinal absorption (HIA)

Reference	Class(es) studied	Dataset[1] size
El-Deeb et al. (2010)	Cyclic arylsulfonylureas	13
Hou et al. (2007b)	Drugs and diverse drug-like molecules	553
Hou et al. (2007c)	Drugs and diverse drug-like molecules	578
Jung et al. (2007)	Oligopeptides (heptapeptide sequences)	852
Verma et al. (2007)	Drugs	57
Subramanian & Kitchen (2006)	Drugs	30
Deconinck et al. (2005)	Drug-like molecules	141+27
Jones et al. (2005)	Drugs	38+131
Liu et al. (2005)	Diverse drugs	113+56
Polley et al. (2005)	Drug-like compounds	NA
Bai et al. (2004)	Diverse drugs	1260
Perez et al. (2004)	Drugs	82+127 & 109
Sun (2004)	Drugs	169
Wegner et al. (2004)	Drugs and drug-like compounds	172+24
Xue et al. (2004b)	P-gp substrates and non-substrates	196
Niwa (2003)	Drugs and drug-like compounds	76+10
Wolohan & Clark (2003)	Drugs and drug-like molecules	86
Zmuidinavicius et al. (2003)	Drug-like compounds	over 1000
Abraham et al. (2002)	Drugs	127
Deretey et al. (2002)	Passively transported drugs	93+31
Klopman et al. (2002)	Drugs	417+50
Raevsky et al. (2002)	Drugs	100
Zhao et al. (2002)	Drugs	238
Agatonovic-Kustrin et al. (2001)	Drugs	76+10
Norinder & Österberg (2001)	Drugs	13+7
Zhao et al. (2001)	Drugs	38+131
Egan et al. (2000)	Drugs and drug-like molecules	234
Ertl et al. (2000)	Drugs	20
Österberg & Norinder (2000)	Drugs	20
Raevsky et al. (2000)	Passively transported drugs	32
Clark (1999a)	Diverse drugs	20+74
Ghuloum et al. (1999)	Diverse molecules	20
Norinder et al. (1999)	Diverse drug-like compounds	13+7
Oprea & Gottfries (1999)	Drugs	NA
Sugawara et al. (1998)	Drugs	NA
Wessel et al. (1998)	Drugs and drug-like compounds	76+10
Palm et al. (1997)	Drugs	20

[1] Training & Test Set (+ Prediction Set, if applicable); NA – information not available

Table 8. Literature models for human oral bioavailability (F)

Reference	Class(es) studied	Dataset[1] size
El-Deeb et al. (2010)	Cyclic arylsulfonylureas	13
Ma et al. (2008)	Structurally diverse drugs	690+76
Moda et al. (2007a)	Structurally diverse molecules	302
Wang et al. (2006)	Structurally diverse drugs	367
Stoner et al. (2004)	Structurally diverse molecules (neutral, basic, acidic)	140
Turner et al. (2004a)	Structurally diverse drugs	152+15
Pintore et al. (2003)	Drugs/Diverse molecules	272/235
Turner et al. (2003a)	Drugs	169
Wolohan & Clark (2003)	Structurally diverse molecules including drugs	198/408
Veber et al. (2002)*	Drug candidates (* Study for rat)	over 1100
Andrews et al. (2000)	Drugs	591
Yoshida & Topliss (2000)	Structurally diverse drugs	232
Hirono et al (1994)	Drugs	188

[1] Training & Test Set (+ Prediction Set, if applicable); NA – information not available

Table 9. Literature models for blood/brain barrier penetration (logBB)

Reference	Class(es) studied	Dataset[1] size
Vilar et al. (2010)	Diverse compounds (* Data for rat mainly) For external validation: CNS active and inactive chemicals	307+1457
Geldenhuys et al. (2010)	Semi-rigid cyclic and acyclic bis-quaternary ammonium analogs	4/5
El-Deeb et al. (2010)	Cyclic arylsulfonylureas	13
Tintori et al. (2009)	Various pyrrolo-pyrimidine c-Src inhibitors	80
Karelson et al. (2008)	Structurally diverse molecules, mainly drugs	60
Konovalov et al. (2008)	Volatile and non-volatile organic compounds and drugs	289
Obrezanova et al. (2008)	Diverse compounds	151+143/292
Van Damme et al. (2008)	Diverse compounds	82
Zhang et al. (2008a)	Drugs and diverse organic compounds	159+99 & 267
Zhang et al. (2008b)	Drugs, small structure-simple molecules, carboxylic acids, alkaloids	160
Al-Fahemi et al. (2007)	Diverse compounds	42
Deconinck et al. (2007)	Drugs	244
Dureja & Madan (2007)	Diverse compounds	62
Konovalov et al. (2007)	Volatile and non-volatile organic compounds and drugs	291
Obrezanova et al. (2007)	Diverse drugs and other small molecules	106
Wichmann et al. (2007)	Diverse neutral molecules	103
Zhao et al. (2007)	BBB crossing and non-crossing drugs, P-gp substrates	1593
Abraham et al. (2006)	Drugs, volatile and non-volatile organic compounds	207/302
Deconinck et al., (2006)	Drugs	147
Dureja & Madan (2006)	Diverse compounds	28
Garg & Verma (2006)	Diverse compounds	191
Gerebtzoff & Seelig (2006)	Diverse compounds	55+43
Katritzky et al. (2006)	Diverse drugs	113
Klon et al. (2006)	Diverse compounds	178
Zhang (2006)	Diverse compounds	80
Li et al. (2005)	Diverse compounds	415
Ma et al. (2005)	Diverse compounds	45
Narayanan & Gunturi (2005)	Diverse compounds	88
Yap & Chen (2005a)	Drugs	155
Zhang (2005)	Neutral and ionized compounds	265
Abraham (2004)	Neutral compounds	30
Cabrera et al. (2004)	Diverse compounds	119+33
Dorronsoro et al. (2004)	Structurally diverse drugs	35
Fu et al. (2004)	Drugs	61
Pan et al. (2004)	Structurally diverse compounds	150
Stanton et al. (2004)	Diverse drugs	47/56
Sun (2004)	Structurally diverse compounds	57
Winkler & Burden (2004)	Diverse drugs and other small molecules	106
Zhang (2004)	Diverse neutral compounds	215
Adenot & Lahana (2004)	BBB crossing and non-crossing drugs, P-gp substrates	1686

Reference	Class(es) studied	Dataset[1] size
Hou & Xu (2003)	Diverse organic compounds	115+37
Hutter (2003)	Diverse compounds	90
Lobell et al. (2003)	Diverse compounds	65
Subramanian & Kitchen (2003)	Structurally diverse compounds	281+181
Wolohan & Clark (2003)	Drugs	55
Atkinson et al. (2002)	Drugs	NA
Doniger et al. (2002)	Diverse classes of molecules, including drugs	324
Hou & Xu (2002)	Structurally diverse compounds	96
Iyer et al. (2002)	Structurally diverse compounds	63
Ooms et al. (2002)	Diverse compounds	79
Rose et al. (2002)	Structurally diverse compounds, including drugs and drug-like molecules	106+28 & 20039
Kaznessis et al (2001)	Drugs	80
Keserü & Molnár (2001)	Diverse compounds	60/85
Klamt et al. (2001)	Diverse compounds	65?
Liu et al. (2001)	Diverse compounds	66
Norinder & Österberg (2001)	Drugs	58
Platts et al. (2001)	Diverse compounds, mainly drugs	148
Crivori et al. (2000)	Diverse compounds, mainly drugs	230
Ertl et al. (2000)	Drugs	45
Feher et al. (2000)	Structurally diverse compounds	75/86
Österberg & Norinder, (2000	Drugs	45/70
Ajay et al. (1999)	Diverse, CNS active and inactive molecules	275
Clark (1999b)	Diverse organic compounds	60/70
Kelder et al. (1999)	Drug molecules	45
Luco (1999)	Structurally diverse compounds	95
Norinder et al. (1998)	Structurally diverse compounds	63
Abraham et al. (1997)	Zwitterionic and non-zwitterionic ampholytes (nitrazepam, albendazole sulfoxide, and sulfadimidine, morphine, difloxacin, and niflumic acid)	6
Salminen et al. (1997)	Drugs	23
Brewster et al. (1996)	Solutes and drugs	60
Kaliszan & Markuszewski (1996)	Diverse compounds	20
Lombardo et al. (1996)	Structurally diverse compounds, from simple solutes to histamine H_2 antagonists	61
Abraham et al. (1995)	Structurally diverse compounds	57
Abraham et al. (1994)	Structurally diverse compounds	57
Calder & Ganellin (1994)	Structure-diverse histamine H_2 receptor antagonists	
van de Waterbeemd & Kansy (1992)	Structure-diverse histamine H_2 receptor antagonists	20
Young et al. (1988)	Structure-diverse histamine H_2 receptor antagonists	6/20

[1] Training & Test Set (+ Prediction Set, if applicable)

Table 10. Literature models for plasma protein binding

Reference	Class(es) studied	Dataset[1] size
El-Deeb et al. (2010)	Cyclic arylsulfonylureas	13
Luan et al. (2009)	Commercially available cephalosporins	28
Karelson et al. (2008)	Drugs	85
Ma et al. (2008)	Structurally diverse drugs	692+161
Weaver & Gleeson (2008)	Diverse molecules	897
Deeb & Hemmateenejad (2007)	Drugs	94
Fu et al. (2007)	Cephalosporins	23
Gleeson (2007)	Diverse molecules	897
Moda et al. (2007b)	Diverse molecules	62
Rodgers et al. (2007b)	Diverse molecules	13-25
Rodgers et al. (2007a)	Diverse molecules	11-84
Estrada et al. (2006)	Drugs	88
Gunturi et al. (2006)	Diverse drugs	94
Hall et al. (2006)	Drugs	200
Wang et al. (2006)	Structurally diverse drugs	266
Votano et al. (2006)	Diverse molecules	1008
Yap & Chen (2005a)	Drugs	75+18
Turner et al. (2004b)	Structurally diverse drug-like compounds	62
Xue et al. (2004a)	Diverse, heterogeneous group of commercially available drugs	94
Yamazaki & Kanaoka (2004)	Diverse pharmaceutical compounds	302+20
Hajduk et al. (2003)	Chemically diverse compounds	889
Hall et al. (2003)	Beta-lactams	115
Lobell & Sivarajah (2003)	Diverse molecules	320
Turner et al. (2003b)	Cephalosporins	20
Valko et al. (2003)	Drugs	52+68
Kratochwil et al. (2002)	Drugs	138
Colmenarejo et al. (2001)	Diverse molecules	95
Saiakhov et al. (2000)	Drugs	154

[1] Training & Test Set (+ Prediction Set, if applicable)

Table 11. Diverse types of literature models for metabolism

Reference	Class(es) studied	Dataset[1] size	Approach[2]
CYP REGIOSELECTIVITY			
Hasegawa et al. (2010)	CYP 3A4 substrates (drug candidates)	61	ML
Sheridan et al. (2007)	Diverse compounds	316/124/92+25	RF
Zhou et al. (2006)	CYP3A4 substrates	227	MetaSite/MD (GLUE)
Lewis et al. (2004)	Drugs	NA	MD/HM
Singh et al. (2003)	Drugs	50	TVM
Zamora et al. (2003)	Drugs	NA	GRID HM
de Groot et al. (2002)	Drugs	16	MD/HM/P/MO
Jones et al. (2002)	Organic compounds	20	QM
Lewis (2002)	Drugs	10	MD/HM
Kuhn et al. (2001)	Drugs (sirolimus (rapamycin) and everolimus)	2	MD/HM/QM
Lewis et al. (2000)	Drugs	12	MD/HM
de Groot et al. (1999a)	Drugs	40	MD/HM/P/MO
de Groot et al. (1999b)	Drugs	40	MD/HM/P/MO
Lewis et al. (1997)	Drugs	NA	MD/HM
de Groot et al. (1996)	Drugs	4	MD/HM
Lewis & Lake (1996)	Drugs	16	MD/HM
Lewis et al. (1996a)	Drugs	NA	MD/HM
Modi et al. (1996)	Drugs	NA	MD/HM
Koymans et al. (1992)	Drugs	NA	MD/HM
Korzekwa et al. (1990)	Organic compounds	NA	QM
CYP SUBSTRATE SPECIFICITY			
Vasanthanathan et al. (2010)	CYP 1A2 substrates	8	LIE
Terfloth et al. (2007)	Drugs and drug analogs	379	SVM
De Graaf et al. (2006)	CYP2D6 substrates	65	MD
Manga et al. (2005)	Drugs	96	FIRM
Yap & Chen (2005b)	CYP3A4, CYP2D6 and CYP2C9 substrates	368/198/144	SVM
Balakin et al. (2004a)	Drug-like molecules	491	KL
Locuson et al. (2004)	Benzbromarone analogs	NA	CoMSIA
Haji-Momenian et al. (2003)	Drugs	24+15	CoMFA
Korolev et al. (2003)	CYP substrates and non-substrates	1008	KL
Lewis et al. (2002)	Drugs	11/6/10/8/8/10/10/10	MLR
Snyder et al. (2002)	Drugs	24	CoMFA/HM
Wang & Halpert (2002)	Drugs	16	P/HM

Reference	Class(es) studied	Dataset[1] size	Approach[2]
De Rienzo et al. (2000)	Drugs	50	HM
Ekins et al. (1999a)	Drugs	38+12	P/PLS
Ekins et al. (1999c)	Drugs	16	P/PLS
RATE AND EXTENT OF METABOLISM			
Hu et al. (2010)	Drug candidates (* Study for mouse, rat and human)	2021 (+ 433) for mouse, 19195 (+ 4886) for rat and 5233 (+ 1193) for human	Classification models (Naïve Bayesian classifier)
Embrechts & Ekins (2007)	MetaDrug Database compounds	317	k-PLS
Balakin et al. (2004b)	MetaDrug Database compounds	83	NN
Crivori et al. (2004)	Pharmacia collection compounds	1800	PLS
Jensen et al. (2003)*	Calcitriol analogs (* Study for rat)	87	PLS
Shen et al. (2003)	Diverse chemicals proprietary to GlaxoSmithKline	631	k-NN
Ekins & Obach, (2000)	Drugs	26/18	P/QSAR
CYP INHIBITION			
Ewing & Feher (2010)	CYP 2D6 and CYP 3A4 inhibitors	6216 (CYP 2D6), 7679 (CYP 3A4)	Various models (e.g. PLS QSAR)
Vasanthanathan et al. (2010)	CYP 1A2 inhibitors	13	LIE
Sushko et al. (2010)	CYP 1A2 inhibitors and non-inhibitors	3621	ASNN
Zhou et al. (2007)	CYP 3A4 inhibitors and non-inhibitors	1699	SVM
Burton et al. (2006)	Diverse compounds	498/306+34/58	RP
Arimoto et al. (2005)	Proprietary compounds	4470	RP/BNN/LR/k-NN/ SVM
Chohan et al. (2005)	Diverse compounds	112	PLS/MLR/CART/BNN
Crivori & Pogessi (2005)	CYP2D6 inhibitors	47	GRIND
Iori et al. (2005)	Flavonoids	NA	MO QSAR
Korhonen et al. (2005)	Diverse compounds including naphthalene, lactone and quinoline derivatives	52	3D-QSAR (CoMFA/GRID/GOLPE)
Kriegl et al. (2005)	Diverse drug-like molecules	780	SVM/PLS-DA
O'Brien & de Groot (2005)	CYP2D6 inhibitors and non-inhibitors	600	ANN/BNN
Vaz et al. (2005)	Aryloxypropanolamines	36	CoMSIA
Yap & Chen (2005b)	CYP3A4, CYP2D6 and CYP2C9 inhibitors	241/180/167	SVM
Afzelius et al. (2004)	Diverse CYP2C9 inhibitors	22	GRID/PLS
Kemp et al. (2004)	National Cancer Institute Database compounds	33	HM/MD (GOLD)
Merkwirth et al. (2004)	Diverse compounds	410	k-NN/SVM/RRM
Asikainen et al. (2003)	Diverse molecules, including naphthalenes	42	CoMFA

Reference	Class(es) studied	Dataset[1] size	Approach[2]
Ekins et al. (2003)	Diverse molecules	1756+98	RP
Susnow & Dixon (2003)	Diverse compounds	100	RP
Wanchana et al. (2003)	Drugs	53	GA-PLS
Afzelius et al. (2002)	Diverse CYP2C9 inhibitors and non-inhibitors	42	GRIND/PLS-DA
Molnar & Keseru (2002)	Genetest Database compoundds	290	ANN
Zuegge et al. (2002)	Diverse molecules	311	PLS/ANN
Afzelius et al. (2001)	Competitive CYP2C9 inhibitors	21+8	3D-QSAR (GRID)/HM
Poso et al. (2001)	CYP2A5 and CYP2A6 inhibitors	28	CoMFA
Ekins et al. (2000a)	Competitive CYP2C9 inhibitors	9/29/13	P(Catalyst)/PLS
Moon et al. (2000)	Flavonoid derivatives	19	MLR/ANN
Rao et al. (2000)	Diverse drugs	14	CoMFA
Ekins et al. (1999b)	Competitive CYP3A4 inhibitors	14/32/22	P(Catalyst)/PLS
Lewis & Lake (1998)	Omeprazole	1	MM/QSAR
Jones et al. (1996b)	CYP2C9 inhibitors	14	CoMFA
Strobl et al. (1993)	CYP2D6 inhibitors	6	CoMFA
PHASE II METABOLISM			
Sorich et al. (2004)	UGT substrates and non substrates	50÷250	SVM/PLS-DA
Smith et al. (2003)	Diverse UGT1A4 substrates	24 (18+6)	P/UFS-PLS/SOMFA
Sorich et al. (2003)	UGT substrates and non substrates	50÷250	PLS-DA/BRANN/SVM
Ethell et al. (2002)	Simple phenols	24/10	QSAR
Sorich et al. (2002	Diverse UGT 1A1 substrates	23 (18+5)	P/2D QSAR/3D QSAR
Cupid et al. (1999)*	Substituted benzoic acids (* Study for rat)	22	QSMR
Cupid et al. (1996)*	Substituted benzoic acids (* Study for rabbit)	24	QSMR
Soffers et al. (1996)	Fluoronitrobenzenes	NA	MO QSAR
Holmes et al. (1995)*	Substituted phenols (* Study for rat)	15	QSMR
Ghauri et al. (1992)*	Substituted benzoic acids (* Study for rat)	14	QSMR
Kim (1991)	Drugs	NA	QSAR

[1] Training and/or Test Set (+ Prediction Set, if applicable); NA – information not available;

[2] ASNN – associative Neural Networks; BNN – Bayesian Neural Networks; BRANN – Bayesian Regularized Artificial Neural Network; CART - Classification And Regression Trees; CoMFA - Comparative Molecular Field Analysis; CoMSIA - Comparative Molecular Similarity Index Analysis; FIRM - Formal Inference-Based Recursive Modelling; HM – Homology Modelling; KL – Kohonen Learning; k-NN – k-Nearest Neighbour; k-PLS - kernel-Partial Least Squares; LIE – Linear Interaction Energy; LR – Logistic Regression; MD – Molecular Docking; ML – Machine Learning; MLR – Multiple Linear Regression; MM – Molecular Modelling; MO – Molecular Orbital calculations; NN – Neural Networks; P – Pharmacophore; PLS – Partial Least Squares; PLS-DA – Partial Least Squares Discriminant Analysis; QM – Semi-Empirical Quantum-Mechanical calculations; QSMR – Quantitative Structure-Metabolism Relationship; RF – Random Forest; RRM - Ridge Regression Modelling; SOMFA – Self-Organizing Molecular Field Analysis; SVM – Support Vector Machine; TVM – Trend Vector Model; UFS - Unsupervised Forward Selection;

Table 12. Literature models for excretion

Reference	Class(es) studied	Dataset[1] size
Paixäo et al. (2010)	Diverse drugs (*In vitro* human hepatic clearance data)	89+112
Doddareddy et al. (2006)	Diverse compounds containing CNS and non-CNS drugs	150
Wang et al. (2006)	Structurally diverse drugs	224
Yap et al. (2006)	Structurally diverse compounds, mainly drugs	503
Ng et al. (2004)	Antimicrobial agents	44
Turner et al. (2004b)	Structurally diverse drug-like compounds	62
Karalis et al. (2003)	Cephalosporins	23
Manga et al. (2003)	Drugs	160
Turner et al. (2003b)	Cephalosporins	20
Agatonovic-Kustrin et al. (2002)	Drugs	123
Cantelli Forti et al. (2002)	Nitroheterocyclic compounds	26
Karalis et al. (2002)	Structurally unrelated drugs	272
Cupid et al. (1996)	Substituted benzoic acids	24
Holmes et al. (1995)	Substituted phenols and their sulphate and glucoronide metabolites	15
Anderson et al. (1992)	Structural analogs of peptide formyl-met-leu-phe	24
Meskin et al. (1985)	Diverse basic and acidic drugs	NA

[1] Training & Test Set (+ Prediction Set, if applicable); NA – information not available

European Commission

EUR 24377 EN – Joint Research Centre – Institute for Health and Consumer Protection

Title: Review of QSAR Models and Software Tools for predicting Biokinetic Properties

Author(s): Mostrag-Szlichtyng A, and Worth A

Luxembourg: Office for Official Publications of the European Communities

2010 – 60 pp. – 21 x 29,7 cm
EUR – Scientific and Technical Research series – ISSN 1018-5593
ISBN 978-92-79-15854-4
DOI 10.2788/94537

Abstract

In the assessment of industrial chemicals, cosmetic ingredients, and active substances in pesticides and biocides, metabolites and degradates are rarely tested for their toxicologcal effects in mammals. In the interests of animal welfare and cost-effectiveness, alternatives to animal testing are needed in the evaluation of these types of chemicals. In this report we review the current status of various types of *in silico* estimation methods for Absorption, Distribution, Metabolism and Excretion (ADME) properties, which are often important in discriminating between the toxicological profiles of parent compounds and their metabolites/degradation products. The review was performed in a broad sense, with emphasis on QSARs and rule-based approaches and their applicability to estimation of oral bioavailability, human intestinal absorption, blood-brain barrier penetration, plasma protein binding, metabolism and. This revealed a vast and rapidly growing literature and a range of software tools.

While it is difficult to give firm conclusions on the applicability of such tools, it is clear that many have been developed with pharmaceutical applications in mind, and as such may not be applicable to other types of chemicals (this would require further research investigation). On the other hand, a range of predictive methodologies have been explored and found promising, so there is merit in pursuing their applicability in the assessment of other types of chemicals and products. Many of the software tools are not transparent in terms of their predictive algorithms or underlying datasets. However, the literature identifies a set of commonly used descriptors that have been found useful in ADME prediction, so further research and model development activities could be based on such studies.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-NA-24377-EN-C

ISBN 978-92-79-15854-4

Publications Office
Publications.eu.int