QSPR in Oral Bioavailability: Specificity or Integrality?

M.Á. Cabrera-Pérez*[,1,2,3], H. Pham-The[1], M. Bermejo[2], I.G. Álvarez[2], M.G. Álvarez[2] and T.M. Garrigues[3]

[1]*Molecular Simulation & Drug Design Group, Central of Chemical Bioactive, Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba;* [2]*Department of Engineering, Area of Pharmacy and Pharmaceutical Technology, Miguel Hernández University, 03550 Sant Joan d'Alacant, Alicante, Spain;* [3]*Department of Pharmacy and Pharmaceutical Technology, University of Valencia, Burjassot 46100, Valencia, Spain*

Abstract: During the last decade the technological advances in drug discovery changed the absorption, distribution, metabolism, excretion and toxicity (ADMET) profiles of New Chemical Entities (NCEs). Among ADMET processes, absorption plays an important role in the research and development of more effective orally administered drugs. Although significant progress has been made in *in vitro*, *in situ* and *in vivo* experimental determinations of absorption, the development of *in silico* methodologies has emerged as a cheaper and fast alternative to predict them. Even though several *in silico* models have been described in the literature to predict oral bioavailability and related properties, the prediction accuracy and their potential use is still limited. The low precision and high variability of data, the lack of a complete experimental and theoretical validation of *in silico* approach, and above all, the multi-factorial nature of the oral absorption term, make the development of predictive *in silico* models a thorny task. The present review discusses several important advances regarding the QSPR approaches used in the development of predictive oral bioavailability models. The importance of fixing the problem associated with data resource, as well as improving the reliability of *in silico* results is highlighted. Optimization of individual properties along the absorption process must be integrated in a multi-objective scenario for studying oral bioavailability behavior in the early drug discovery and development.

Keywords: ADMET, QSPR, oral bioavailability, oral absorption, intestinal permeability, drug solubility, CYP3A4 metabolism, P-gp efflux.

INTRODUCTION

The process for a new chemical entity (NCEs) to become a drug candidate and then reach the market usually involves hundreds of millions of dollars of investment and more than a decade of research. During this time there are many different events, which can halt this development. One of the main reasons that prevent a new candidate from becoming a prescription drug is a poor pharmacokinetic profile. In the 90s, some studies showed that ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) problems were responsible for 40% of drug failure, but this value decreased substantially to 10% in 2000 [1-3].

On the other hand the use of combinatorial methods during the last 20 years has brought about a vast number of more lipophilic, less soluble and higher molecular weight NCEs than conventional drug entities. These properties are often translated into unfavorable absorption, leading to poor and erratic bioavailability [4]. The early identification of drug bioavailability can assist researchers not only to a better selection of candidates for further development by rejecting those with lower chance of success [5].

*Address correspondence to this author at the Molecular Simulation & Drug Design Group, Central of Chemical Bioactive, Central University of Las Villas, Santa Clara, 54830, Villa Clara, Cuba; Tel: 53-42-281192; Fax: 53-42-281130; E-mails: macabrera@uclv.edu.cu, migue@gammu.com, macabreraster@gmail.com

Even with the advantages of traditional *in vivo* and *in vitro* methods [6, 7], the experimental measurement of the absorption properties is still expensive and time consuming. In this context, some *in vitro* methods have been introduced to predict *in vivo* behavior [8]. Nowadays, these *in vitro* methods can no longer match the demand in throughput as the number of compounds that can be generated has increased dramatically [9] and alternative *in silico* tools have emerged to predict ADMET properties [10, 11], even before synthesis.

Several computational methods to predict oral absorption have been developed. Most of them are focused on quantitative structure-property relationship (QSPR) studies, using different types of molecular descriptors and statistical methodologies. Although the *in silico* approach has some important advantages compared to the experimental methodologies, the prediction of complex biological properties such as bioavailability is a big challenge due to multiple factors involved such as gastrointestinal transit, chemical stability in the gastrointestinal tract, intestinal permeability and first-pass effect of gut wall and liver metabolism [12].

This review article will discuss QSPR approaches used in the development of predictive oral drug bioavailability models and related absorption properties, their advantages, limitations and future challenges in the discovery and development of drugs.

1875-5607/12 $58.00+.00

1. GENERALITIES ABOUT ORAL BIOAVAILABILITY

According to the Food and Drug Administration (FDA) in the United States and the European Medicine Agency (EMA) drug bioavailability is the rate and extent to which the active substance or active moiety is absorbed from a pharmaceutical form and becomes available at the *site of action* [13, 14]. Based on the unfeasibility of measuring a drug's level at the site of action, the useful definition of bioavailability is indeed based on plasma concentrations [4].

The bioavailability F of an orally administered dose is comprised of the individual fractions that overcome the various barriers encountered by the drug during its way through the gut lumen and through the liver. Mathematically, it can be expressed as a direct relationship, as in the following equation:

$$F = F_a \times F_g \times F_h \quad (1)$$

where, Fa represents the intestinal absorption fraction, Fg stands for the intestinal first-pass fraction ($E_g = 1 - F_g$), and Fh is the hepatic extraction ($E_h = 1 - F_h$).

The experimental approach to estimate oral bioavailability in human volunteers or in animal models involves the administration of the drug by the intravenous route (iv) and the oral one. Plasma levels are measured at the adequate sampling points and the area under the curve (AUC) is calculated by the trapezoidal method for both administrations. It can be demonstrated that AUC_0^∞ (from time zero to infinity) corresponds to:

$$AUC_0^\infty = \frac{F * D}{Cl} \quad (2)$$

where, D is the dose administered and Cl the drug clearance. Taking into account that the availability of the drug from the iv administration is complete (100%), systemic bioavailability (in extent) F_{sys} from the oral administration can be calculated from the following expression,

$$F_{sys} = \frac{AUC_{oral} / D_{oral}}{AUC_{iv} / D_{iv}} \quad (3)$$

The basic assumption in this expression is that clearance is the same for both administrations.

If the drug has a nonlinear profile, the clearance is not constant and therefore F cannot be estimated using the above equation.

In some other instances, intravenous administration is not feasible for different reasons. In such cases, the relative bioavailability is calculated by comparing the pharmaceutical form to another standard (i.e. oral solution):

$$F_{rel} = \frac{AUC_{oral}}{AUC_{oralsolution(standard)}} \quad (4)$$

As the concept of bioavailability includes not only extent but also rate of access to the systemic circulation, parameters to compare rate are needed. Regulatory bodies accept the maximal plasma level C_{max}, and the time at which this level is reached t_{max}.

As can be inferred, the bioavailability studies depend on drug blood sampling. Standard procedures are set up for human volunteers and animal models of enough size and weight as dogs, monkeys etc. In smaller animal models, as rat or mice, it is usually necessary to permanently cannulate the jugular or femoral vein or the use of other sampling methods (retro-orbicular, tail, etc) with the consequent limitations of the number of samples per animal. In this latter case the use of the adequate mathematical procedures for the AUC estimation are warranted [15] .

In summary, oral bioavailability is a complex property that is influenced by many physicochemical and biological factors (See Fig. **1**). Although all these factors are known, their interaction is not always completely clear and only few of them have been used to predict bioavailability [16, 17]. In this context, reasonably successful *in silico* models have been developed for solubility, intestinal permeability, human intestinal absorption (HIA) and first-pass effects (P-glycoprotein, P-gp and Cytochrome P450, CYP). Some of the most relevant results will be described in this review in order to highlight their potential use in drug design and early drug discovery.

2. ORAL ABSORPTION PREDICTION

Taking into consideration that the F value is the result of complex biological mechanisms, the most reasonable sequence in oral bioavailability studies should be: (1) assessment of the physical diffusion and active transport through the membrane followed by (2) prediction of subsequent intestinal and hepatic phase-I metabolism. The first step is related to physical-chemical properties of the drug (stability, pKa, molecular size and solubility), the type of transport across the intestinal membrane (passive or active uptake) and efflux of compounds across the cell membrane (e.g. P-gp and multidrug resistance associated proteins, MRPs). The second step is concerned with the determination of intestinal and hepatic metabolism mainly produced by enzymes of the sub-family of CYP [18], with the CYP3A4 enzyme being the major determinant of oral bioavailability [19, 20].

One of the main challenges for oral drug administration is to achieve a maximum absorption through the intestinal tract with low intestinal first-pass metabolism. The intestinal absorption fraction of a drug (Fa) depends basically on its solubility and gastrointestinal permeability. Both properties integrate the scientific framework of the Biopharmaceutical Classification System (BCS) [21, 22].

In drug discovery, therefore, it would be highly desirable to be able to separate and quantify distinctly luminal events from systemic ones and permeability issues from first-pass effects, for individual compounds under investigation [23]. Almost all reported studies have dealt with the problem of using different computational approaches for prediction of individual properties such as solubility, intestinal

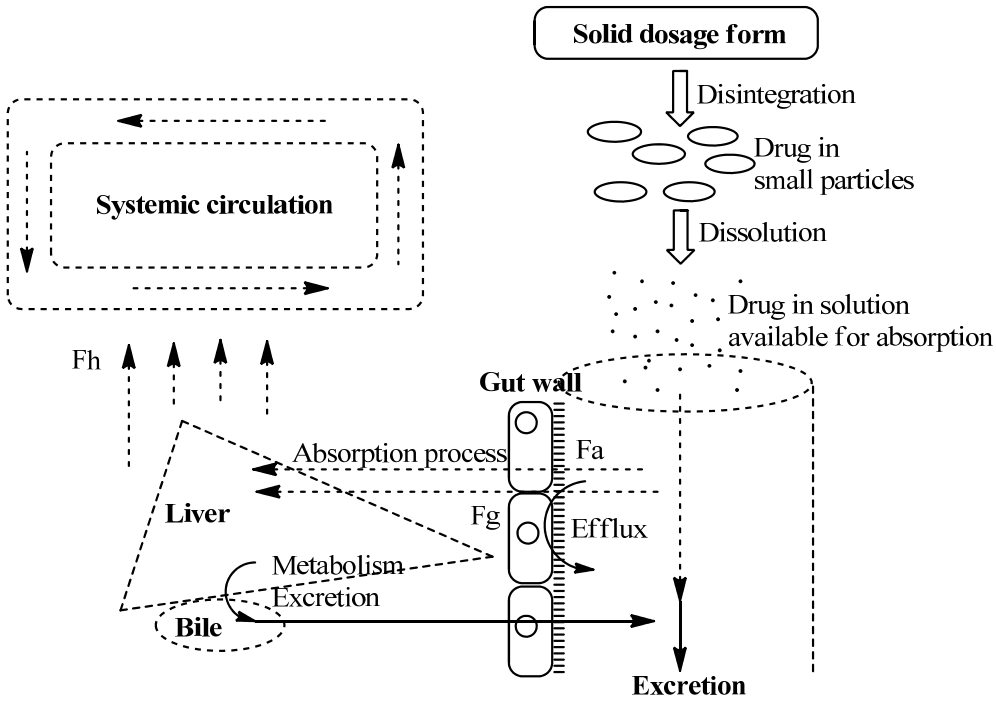

Fig. (1). Schematic representation of oral delivery process of drugs to systemic circulation, which displayed the three pharmacokinetic processes (dissolution, permeation though intestinal membrane and first-pass effects herein are efflux and human liver biotransformation) of drug via oral absorption. Also this schematic depicted the three main factors defined oral bioavailability (Fa, Fg and Fh) as well as the general relationship provided by the Fick's law (See Eq. **1**).

permeability, human intestinal absorption and the first-pass metabolism effects. A summary of the main QSPR models for these individual properties, their advantages and limitations in the prediction of oral bioavailability will be discussed below.

2.1. SOLUBILITY

The solubility of any compound can be defined as the amount of substance that has passed into solution when the equilibrium between the solution and the excess of undissolved substance at a given temperature and pressure is attained. Taking into consideration that compounds with similar values of aqueous solubility could be administered at very different doses, a dimensionless parameter called *Dose Number* (D_0) is used to classify drug solubility [24]. This parameter is defined as the ratio of drug concentration in the administered volume to its saturation solubility in water.

Ideally, any computational model to predict the effect of solubility on absorption should be based on the concept of D_0 to better characterize the drug solubility. However, few works have studied this parameter quite deeply. Many *in silico* models were developed to predict solubility of non drug-like molecules, considering intrinsic solubility as experimental data. Although these datasets were comprised of molecules with limited structural complexity with respect to functional groups and ring systems, the computational models obtained have adequate accuracy and predictive capacity [25]. Nevertheless, these models have little applicability to drug solubility due to the unknown or high experimental error, the lack of structural diversity, the non-consideration of salt and/or common ion issues, the crystal packing effects and the poor pharmaceutical relevance or inadequate range of solubility [26]. On the contrary, several authors obtained models with high level of performance and prediction using various small drug-like datasets [27-29].

In order to avoid the limited structural diversity of some databases, Delaney [30], Votano *et al*. [31] and Wang *et al*. [32] developed good statistical models using large datasets (> 2500 compounds) but their predictive capacity should be consistently evaluated with large external sets. Recently, Johnson *et al*. [33] published a QSPR model for the prediction of aqueous solubility that includes crystal packing, intrinsic solubility, and specific ionization effects. This approach suggested a novel method able to address the issues related to solubility prediction.

Although many efforts have been made to develop good computational models for solubility prediction of drug-like molecules, the current methods have small datasets and a limited predictive applicability compared with non drug-like molecule models. Beyond these differences, we consider that both types of models are necessary during the drug discovery process because of the significant impact of solubility on the drug uptake, transport and eventually bioavailability. Models for solubility prediction of non drug-like molecules could be used during the stage of hit identification while models for drug-like molecules would be used during the lead optimization process.

Table 1. Different *in silico* Models to Predict Solubility

Reference	Modeling Methods	Descriptors	Dataset	Performance of the Best Model
Huuskonen *et al.*[97]	ANN[a]	Atom-type E-state, topological	$N^{i}_{tr+test}$=1297	$Q^{1}_{Training}$=94% Q^{2}_{Test}=88%
			$N^{ii}_{ext-set}$=21	$Q^{3}_{External\ set}$=91%
McElroy and Jurs [98]	MLR[b] ANN	Topological, geometrics, electronic and hybrid descriptors	$N_{tr+test}$=399	$Q_{Training}$=90% Q_{Test}=88% $Q_{External\ set}$=53%
Tetko *et al.*[99]	ANN	Molecular weight, atom-type E-state	$N_{tr+test}$=1291	$Q_{Training}$=94% Q_{Test}=91%
			$N_{ext-set}$=21	$Q_{External\ set}$=90%
Liu and So [100]	ANN	Hydrophilicity, molecular weight and topological	$N_{tr+test}$=1312	$Q_{Training}$=93% Q_{Test}=93%
			$N_{ext-set}$=21	$Q_{External\ set}$=89%
Livingstone *et al.* [101]	MLR ANN CCA[c]	Electrotopological descriptors	$N_{tr+test}$=688	$Q_{Training}$=90% $Q_{validation}$=92% $Q_{Test\ set}$=84%
			$N_{ext-set1}$=19 $N_{ext-set2}$=21	$Q_{External\ set\ 1}$=91% $Q_{External\ set\ 2}$=86%
Cheng *et al.* [102]	GA[d]/MLR	Topological, geometrics and electronic	$N_{tr+test}$=321	$Q_{Training}$=95% Q_{Test}=84%
Gao *et al.* [103]	GA/MLR	MOE descriptors	$N_{tr+test}$=1179	$Q_{Training}$=91% Q_{Test}=91%
Engkvist and Wrede [104]	ANN	Topological and constitutional descriptors	$N_{tr+test}$=3351	$Q_{Training}$=91% Q_{Test}=89%
			$N_{ext-set}$=307	$Q_{External\ set}$=86%
Yan and Gasteiger [105]	ANN	Topological descriptors	$N_{tr+test}$=1293	$Q_{Training}$=92% Q_{Test}=94%
Yan and Gasteiger [106]	ANN	3D descriptors	$N_{tr+test}$=1293	$Q_{Training}$=93% Q_{Test}=92%
Cheng and Merz [107]	GA/MLR	Constitutional, lipophilicity, topological, atom-type E-state and Jurs's charged partial surface area parameters	$N_{tr+test}$=775	$Q_{Training}$=84%
Butina and Gola [108]	PLS[e] Cubist method	Molecular properties, lipophilicity	$N_{tr+test}$=3328	$Q_{Training}$=80% Q_{Test}=74%
			$N_{ext-set}$=11	$Q_{External\ set}$=89%
Manallack *et al.* [109]	ANN	3D BCUT descriptors	$N_{tr+test}$=788	$Q_{Training}$=87.18%
Schaper *et al.* [110]	MLR	HYBOT descriptors	$N_{tr+test}$=787	$Q_{Training}$=94% Q_{Test}=92%

(Table 1). Contd…..

Reference	Modeling Methods	Descriptors	Dataset	Performance of the Best Model
Wegner and Zell [111]	ANN	Topological descriptors, electronic descriptors	$N_{tr+test}$=1269	$Q_{Training}$=94% Q_{Test}=82%
Votano *et al*. [31]	MLR ANN	Constitutional descriptors, topological descriptors	$N_{tr+test}$=3343	$Q_{Training}$=88% Q_{Test}=77%
Delaney [30]	MLR	Physicochemical descriptors	$N_{tr+test}$=2874	$Q_{Training}$=72%
			$N_{ext-set}$=21	$Q_{External\ set}$=85%
Hou *et al*. [112]	MLR	Atom types	$N_{tr+test}$=1290	$Q_{Training}$=96%
			$N_{ext-set}$ =21	$Q_{Test\ 1}$=94%
Wang *et al*. [113]	PLS	3D Sybyl 7.0 descriptors, physicochemical descriptors	N^{iii}_{total}=1708	$Q^4_{ASMS\ model}$=87.2% $Q^5_{ASMS-LOGP}$=88.6%
Huuskonen *et al*. [29]	MLR	Physicochemical descriptors	N_{total}=191 (drug-like compounds)	$Q_{Training}$=87% $Q_{External\ set\ 1}$=80% $Q_{External\ set\ 2}$=88%
Wang *et al*. [32]	MLR	Atom Type Count, 3D Surface Area (SAS), physicochemical descriptors	N_{total}=3664	$Q^6_{ASM-ATC-LOGP}$=83.2% $Q^7_{ASM-ATC\ Model}$=82.1% $Q^8_{ASM-SAS-LOGP}$=82.7% $Q^9_{ASM-SAS}$=81.8%

[a]Artificial neural network; [b]multiple linear regression; [c]canonical correlation analysis; [d]genetic algorithm; [e]partial least square; [i]number of compounds used for training and test set; [ii]number of compounds used for external set; [iii]total number of compound used in study; [1,2]overall accuracy of correlation or classification of training set and/or test set; [3]accuracy of prediction of external set; [4-9]accuracy of aqueous solubility model based on specified descriptors such as atom type counts, ClogP and solvent accessible surface areas.

2.2. PERMEABILITY

Effective permeability (Peff) is one of the parameters used to measure rate and extent of intestinal absorption (Fa). It is commonly reported as an apparent permeability coefficient (Papp or Pm) using different cell culture models and is usually reported in cm/s. The most commonly used model is the Caco-2 cell which is a human colon carcinoma cell line [34]. It is generally accepted that good permeability through Caco-2 monolayers is a reliable index of good *in vivo* absorption [35, 36], unless dissolution is a problem. This *in vitro* model has been recommended by the US FDA for determination of permeability of compounds to be classified according to the BCS [37]. Nevertheless, one of the main problems of this *in vitro* assay is the considerable inter- and intra-laboratory variability [38, 39] which makes it difficult to combine Caco-2 permeability data from different sources to form one large dataset [40]. Several QSPR models to predict Caco-2 permeability have been published elsewhere (See Table **2**).

As can be seen in Table **2**, different statistical methodologies and datasets have been used to develop computational models of Caco-2 permeability. Although very good regression models have been obtained, we consider that the classification model is the best option taking into account the great variability in the experimental determination of Caco-2 permeability. A good strategy for the future development of useful *in silico* Caco-2 permeability models is to carry out classification approaches using the permeability ranges of compounds with high intestinal absorption such as Metoprolol (Fa = 95; Papp > 20×10^{-6} cm/s). A possible classification scheme would be: Low permeability: Papp < 2×10^{-6} cm/s; Moderate permeability: 2×10^{-6} < Papp < 20×10^{-6} cm/s and High permeability: Papp > 20×10^{-6} cm/s [41, 42]. On the other hand, the potential of these *in silico* models in the prediction of human intestinal permeability and their application along with *in silico* solubility models is a relevant option to select drug candidates with good oral absorption during the early drug discovery.

The BCS is a scientific framework to understand drug absorption in terms of *in vitro* aqueous solubility and intestinal permeability [43]. However, as was previously discussed, there has been a considerable research over the last decade on *in silico* methods for prediction either solubility or intestinal permeability separately. In this context, computational classification models to predict the BCS/BDDCS (Biopharmaceutical Drug Disposition Classification System) class from molecular structure without running expensive and time-consuming *in vitro* and *in vivo* permeability studies are outstanding [44-47]. Even though the results obtained to predict absorption are adequate, in our opinion in the future each class of the

Table 2. Different *in Silico* Models to Predict Caco-2 Permeability

Reference	Statistical Method	Descriptors	Database	Performance of the Best Model
Fujiwara *et al*. [114]	MLR[a] and ANN[b]	Dipole moment, polarizability and atoms of N, O, and H	N^{i}_{total}= 129	$Q^{1}_{Training}$=58%
Yamashita *et al*. [115]	GA[c]-PLS[d]	Molconn-Z descriptors	N_{total}=73	$Q_{Training}$= 89% Q^{3}_{cv} (LOO) = 83%
Hou *et al.* [116]	MLR	Simple molecular properties	N^{ii}_{tr} = 77 N^{iii}_{test}= 23	$Q_{Training}$= 82% Q^{2}_{Test}=78%
Marrero *et al.*[117]	LDA[e]	Atom-based quadratic indices	N_{tr}=134 N_{test}= 12	$Q_{Training}$=90.3% Q_{Test}=83.3%
Refsgaard *et al.*[118]	Nearest-Neighbor Classification	mLogPa, MWb, HDc, HAd, ROTe, cLogPf, VOLg, SURFh, PSAi	380 (Class 0) 332 (Class 1) $N^{iv}_{ext-set}$=112	Q^{4}_{Class0} =75.5% Q^{5}_{Class1} = 88.5% $Q^{6}_{External\ set}$=84.6%
Guangli and Yiyu [119]	SVM[f]	Chemistry Development Kit (CDK) descriptors	N_{tr} = 77 N_{test}= 23	$Q_{Training}$=88% Q_{Test}=85%
Di Fenza *et al.* [120]	GA-ANN	Volsurf descriptors	N_{tr-1}= 106 N_{test-1} = 50 N_{tr-2}= 101 N_{test-2}= 50	$Q_{Training1}$=72% Q_{cv1} (LOO)= 40% $Q_{Training2}$=75% Q_{cv2} (LOO)= 61%
Castillo-Garit, *et al*. [121]	LDA MLR	Atom-based non-stochastic Linear Indices	N_{tr}=138 N_{test} = 19 N_{tr} = 77 N_{test} = 23	$Q_{Training}$= 90.6% Q_{Test}= 84.2% $Q_{Training}$= 85% Q_{Test}=71%
Paixao *et al*. [122]	ANN	ALOGPS 2.1 program and e-DRAGON 1.0 descriptors	N_{tr} = 192 N_{test} = 59 $N_{ext-set}$=45	$Q_{Training}$= 84% Q_{Test}= 70% $Q_{External\ set}$=77%
Pham The *et al*. [42]	LDA	DRAGON Version 5.4 descriptors	N_{tr} =537 N_{test} =137 $N_{ext-set}$ = 10	$Q_{Training}$= 81.6% Q_{Test}=83.9% $Q_{External\ set}$=80.0%

[a]Multiple linear regression; [b]artificial neural network; [c]genetic algorithm; [d]partial least square; [e]linear discriminant analysis; [f]support vector machine; [i]total number of compound used in study; [ii]Number of compounds used for training set; [iii]number of compounds used for test set; [vi]number of compounds used for external set; [1,2]overall accuracy of correlation or classification of training set and/or test set; [3]accuracy of leave one out cross-validation; [4,5]accuracy of prediction of classification models for class 0 and class 1; [6]accuracy of prediction of external set.

BCS/BDDCS could be better predicted by means of *in silico* models with information of permeability, solubility, P-gp efflux and metabolism, in order to provide useful additional insight into bioavailability properties of drugs.

2.3. Human Intestinal Absorption

The first to attain a high oral bioavailability is to achieve a good oral absorption [48]. The human intestinal absorption (HIA, % oral absorption) is an *in vivo* measure defined as the % dose of orally administered drug to reach the hepatic portal vein [49]. It can also be determined as the % of urinary excretion of drug-related material following oral administration, or the ratio of the total mass absorbed divided by the drug dose (% fractional absorption, Fa). Oral absorption takes into consideration metabolism that occurs in the gut wall, but not first-pass metabolism in the liver.

Considering the important role of HIA, the development of QSPR models can be very useful to speed up the design of new compounds with appropriate HIA profiles, and consequently to reduce time and cost in drug discovery and developmental processes. This issue has been addressed by many authors that reported predictive models for HIA based on different molecular descriptors and statistical methods (See Table **3**).

Analysis of Table **3** shows that, similar to permeability equations, the classification models show better results in the prediction of HIA than numerical ones. Even though large

Table 3. Different *in Silico* Models to Predict Human Intestinal Absorption

Reference	Statistical Method	Descriptors	Database	Performance of the Best Model
Zhao *et al.* [123]	MLR[a]	Abraham descriptors	N^{i}_{tr}= 31 N^{ii}_{test}=138	$Q^{1}_{Training}$ = 85% Q^{3}_{CV} = 78%
Klopman *et al.*[124]	MLR	Substructural molecular descriptors	N_{tr}=417 N_{test}= 50	$Q_{Training}$ = 79% Q^{2}_{Test} = 79%
Cabrera-Pérez *et al.*[125]	LDA[b]	TOPS-MODE descriptors	N_{tr}= 82 N_{test}= 127	89 % of good classification 93 % of good classification
Deconinck *et al.* [126]	Classification and regression trees (CART)	DRAGON descriptors	N_{tr}=141 N_{test}=27	$Q^{4}_{overall}$ = 77.8%-88.9%
Liu *et al.*[127]	Heurist method (HM) and SVM[c]	CODESSA descriptors	N_{tr}=113 N_{test}=56	$Q_{Training}$ = 78% (HM) $Q_{Training}$ = 86% (SVM) Q_{Test} = 70% (HM) Q_{Test} = 73% (SVM)
Hou[128]	-	Physicochemical and topological descriptors	N_{tr}= 480 N_{test} = 98	$Q_{Training}$ = 97.8% (HIA-) $Q_{Training}$ = 94.5% (HIA+) Q_{Test} = 100% (HIA-) Q_{Test} = 97.8% HIA+
Gunturi *et al.*[129]	k-Nearest neighbors (kNN) method along with genetic algorithms (kNN-QSAR-GA)	Structural, physico-chemical, geometrical and topological descriptors	$N^{iii}_{tr/test}$=126/49 $N_{tr/test}$=117/58 $N_{tr/test}$=117/58 $N_{tr/test}$=49/126	Q^{5}_{ext}= 80% Q_{ext}= 70% Q_{ext}= 71% Q_{ext}= 70%
Hou *et al.*[48]	GFA and the multivariate adaptive regression splines (MARS) Recursive partitioning (RP)	TPSA,MW, *N*rot, *N*HBD, *N*HBA, log *P*, log *D*, log *S*, MV, MR, *N*rule-of-5, etc. (*W*), and Zagreb index (Zagreb).	N_{tr} = 455 N_{test} = 98 N_{tr} = 481 N_{test} = 98	$Q_{Training}$ = 84% Q_{Test} = 90% 96% of good classification 96.9% of good classification
Iyer *et al.*[130]	GA-MLR; MLR	MI-QSAR descriptors	N_{tr} =106 N_{test} =21	$Q_{Training}$ = 82% Q_{Test} = 70%
Yan *et al.* [131]	PLS[d] and SVM	ADRIANA code, Cerius index	N_{tr} = 380 N_{test} = 172	$Q_{Training}$ = 72-81% Q_{Test} = 83-89%
Reynolds *et al.*[132]	ADME boxes or algorithm builder descriptors	logPo/w, NHD, Vx, ion fractions (pKa function)	N_{tr}=567 $N1_{test}$=25 $N2_{test}$=22	$Q_{Training}$ = 93% Q_{Test1} = 72% Q_{Test2} = 84%

[a]Multiple linear regression; [b]linear discriminant analysis; [c]support vector machine; [d]partial least square; [i]Number of compounds used for training set; [ii]number of compounds used for test set; [iii]number of compounds used for training/test set; [1,2]overall accuracy of correlation or classification of training set and/or test set; [3]accuracy of prediction of model cross-validation; [4]overall accuracy of classification models for training and test set; [5]accuracy of prediction of external set.

datasets have been used, almost all of them are usually collected from drugs or drug candidates in clinical trials. These datasets may show significant variability from one source to another and often are heavily biased towards compounds with high intestinal absorption values because most of the compounds reported are commercially available drugs [48]. This fact will influence the predictive capacity of the *in silico* models and better predictions will be obtained for compounds with high intestinal absorption values compared to the rest of the dataset.

Another problem is the model capacity to predict the intestinal absorption of compounds with active transport since most of the models were obtained under assumption that compounds of dataset followed a passive absorption transport [25].

Basically, all descriptors used in QSAR/QSPR analysis can be employed in the prediction of intestinal absorption. However, those molecular descriptors related with physicochemical properties are most relevant and their use will allow to gain some insight into factors that are likely to

govern the absorption of drugs and to understand which interactions play an important role during the absorption process.

As oral drug absorption is particularly complex, some researchers have focused their attention in developing rule of-thumb alerts and classification systems to easily identify potential absorption problems. The most cited of these alerts is the 'rule of five' proposed by Lipinski and co-workers, that has come to be a compass for the drug discovery industry [50]. These authors, after analyzing the physicochemical profiles of 2245 orally active drugs from the World Drug Index (WDI) [51], concluded that poor absorption or permeation is more likely when two of the following conditions are fulfilled: molecular weight (MW) > 500; H-bond acceptors (HBA) n > 10; H-bond donors (HBD) n > 5 and calculated log P (ClogP) > 5. A modification of this rule was proposed by Congreve *et al*. [52], which assumed other cutoffs for physicochemical properties [MW ≤ 300, HBA ≤ 3, HBD ≤ 3, ClogP ≤ 3, number of rotatable bonds (RTB) ≤ 3, polar surface area PSA ≤ 60] to design fragment libraries for fragment - based lead generation.

After the 'rule of five', many attempts were made to generate rules-of-thumb to identify well-absorbed candidates. Veber *et al*. proposed a set of rules based on rat oral bioavailability data which states that either (i) PSA ≤140Å^2 and number of rotatable bond ≤10 or (ii) sum of H-bond donors and acceptors ≤ 12 and number of rotatable bond count ≤10 are efficient and selective criteria to discover orally available drugs of high molecular weight [53]. Similar results were confirmed by Lu *et al*. when studied a dataset of 434 molecules with values of oral bioavailability in rats [54]. In 2007, Hou *et al*. examined Veber's rule on a dataset of 773 compounds and demonstrated that these simple rules based on rat oral bioavailability data can not be used to predict adequately human oral bioavailability [55].

On the other hand, the rules of thumb like "rule of five" should be used with care to avoid the possible exclusion of promising compounds [51]. Recently, Giménez *et al*. evaluated around 60 blockbuster drugs (promising marketing drugs) and demonstrated that about 89% of successful drugs were out of the thresholds of the "rule of five" [56]. Other results confirm this statement, for example a study of 1204 US FDA approved small-molecule drugs revealed that only 885 drugs (73%) passed the "rule of five" and 70% of them (619 drugs) are used orally [57]; the 68.7% of the compounds in ACDSD (ACD Screening Database, 2.4 million compounds) and 55% of the compounds in ACD (240 thousand compounds) have no violation of the rule of five [58]. This figures point out that "rule of five" only provides a basic orientation about the possibilities of a molecule to be drug-like.

3. INTESTINAL AND HEPATIC PHASE-1 METABOLISM PREDICTION

Oral bioavailability involves complex biological mechanisms that may reduce the fraction absorbed, i.e. efflux and/or first-pass metabolism. It is well known that both CYP3A isoenzymes and efflux proteins such as ABCB1/ P-gp are major determinants of oral bioavailability [19, 20]. The structural identification of CYPs and P-gp substrates, inhibitors or inducer agents has become an attractive field of QSPR application to improve the predictability on drug´s bioavailability. Many authors have suggested that CYP3A4 and P-gp act in a coordinated manner. This theory is based on their location in the small intestine enterocytes and the overlapping in their specificities for substrates. Nevertheless, no correlation between the tissue levels of CYP3A and P-gp has been found in human enterocytes [59] and the same occurs if the concentration of CYP3A in the intestine is compared with the enzyme levels in the liver [60].

In the following section we will focus in some important approaches to QSPR modeling of intestinal and hepatic first-pass effects and the main achievements and limits of existing strategies.

3.1. Cytochrome P450 (CYP) 3A4

CYP are the main enzymes involved in the biotransformation of drugs and other xenobiotics. They comprise a superfamily of hemeproteins that is subdivided into 18 families and 43 subfamilies. However, only three main P-450 gene families, CYP1, CYP2, and CYP3 are thought to be responsible for drug metabolism. Among these, CYP3A (CYP3A4 and 3A5) and CYP2C (CYP2C8, 2C9, 2C18 and 2C19) are the most abundant subfamilies, accounting for 30% and 20% of total CYP, respectively [61]. In adults, CYP3A4 is the dominant CYP3A isoform in the human small intestine and liver, which is present in more than 60% of the total CYP of the liver and often more than 70% of the enterocyte [62]. Although CYP3A4 preferentially catalyzes the oxidation of lipophilic neutral or basic compounds, its hydrophobic active site is capable of accommodating a wide range of structures, from simple and rigid steroids to macromolecules [63]. CYP3A4 is involved in the metabolism of approximately half of the drugs which are used nowadays [61]. Thus CYP3A4 inhibition should be well attractive for oral drug bioavailability enhancement, especially for carcinogens (cancer chemoprevention strategy). However, some drawbacks such as the variations in the catalytic activity and the limited experimental data restrict the general applicability of QSPR models.

Different computational methodologies such as three-dimensional quantitative structure–metabolism relationships (3D-QSMRs) [64-67], mechanistic models [68, 69] and MetaSite [70] have been used to model CYP3A4, being the QSPR model for 3A4 one of the better choice. Some of the more relevant models to classify CYP3A4 substrates and/or inhibitors are depicted in Table **4**.

QSPR models for predicting the site of compound metabolism have undergone significant advances during the past years [71, 72] (See Table **4**). Three main issues (data, model and prediction) arise with the development and prediction of metabolism by QSPR models.

For the first issue the main problems are related with data reproducibility due to the limited size of datasets and different threshold inputs in the kinetic parameters such as IC_{50}, V_{max}, K_{max} and intrinsic clearance (Cl_{int}) which makes it

Table 4. ***In-Silico*** **Models for CYP3A4-Mediated Drug Metabolism Prediction**

Authors	Methods	Descriptors	Dataset	Overall Accuracy	Metabolized Agents
Molnár and Keserü [133]	ANN[a]	2D descriptors	Genetest database	90%	Inhibitors
Ekins *et al.*[134]	RP[b] (tree)	2500 Chemtree atom descriptors	1756	82%	Inhibitors
Korolev *et al.* [135]	Kohonen self-organizing maps (SOM)	26 physicochemical descriptors	2200	76.7% 62.7%	Substrates Products
Merkwirth *et al.*[136]	Ensemble k-NN[c], SVM[d], and ridge regression	-	410	88%	Low inhibitors High inhibitors
Balakin *et al.*[137]	ANN	Physicochemical descriptors, sum of the squares of vertex valencies	491	91% 97%	High K^{l}_{m} Low K_m
Yap and Chen[138]	Consensus SVM	Dragon descriptors	609	96% 95%	Inhibitors Substrates
Kriegl *et al.*[139]	SVM classification	2-D descriptors VolSurf descriptors AM1 quantum indices	1345	76% 67% 77%	strong inhibitors medium inhibitors weak inhibitors
	SVM regression			51%	
Yap *et al.*[140]	SVM LR[e] LDA[f] PLS[g] C4.5 DT[h] *k*-NN P-NN[i]	Dragon descriptors	-	94.6% 87.8% 48.6% 90.5% 87.8% 98.6% 87.8%	Inhibitors Substrates
Jones *et al.*[141]	*Pharmacophore method**	Lipophilicity, molecular Weight	54	-	Substrates Unknown-substrates
	RP (tree)			75-95%	
	LR			92%	
	Regression tree			<65%	
	LMR[k]			<40%	
Terfloth *et al.* [142]	LR decision tree SVM	Molecular properties, topological descriptors	379 233 external set	90% 83%	Substrates
Feher and Ewing [73]	PLS	Physicochemical descriptors, fraction positive and negative ionization at pH 7.4	4224	80-85%	Inhibitors

[a]Artificial neural network; [b]recursive-partitioning (tree); [c]k-nearest neighbor; [d]support vector machine; [e]logistic regression; [f]linear discriminant analysis; [g]partial least square; [h]C4.5 algorithm and multivariate decision trees; [i]polynomial neural network; [k]multiple linear regression; [l]Michaelis-Menten constant; *pharmacophore method capable to predict metabolic intermediate complex (MIC) formation with CYP3A4 (+b5).

difficult to compare different models and impossible to combine datasets [73].

Concerning the model issue, it is important to remark that drug metabolism is an extremely complex pharmacokinetic process being very difficult to obtain accurate modeling of the drug-metabolic enzyme interactions. Not one modeling technique is consistently better than the others; for this reason, the use of combined models could improve the metabolism prediction (see Table **4**). Some output combinations of QSPR models with pharmacophore-based

Table 5. ***In Silico*** **Models for P-gp Substrate/Inhibitor Prediction**

Authors	Methods	Descriptors	Number of Compounds	Overall Accuracy	Interaction
Bain *et al.*[143]	Classification	Number of rings, molecular weight (MW), logP, hydrogen-bond (H-Bond)	N^{i} = 44	Q^{1}_{subs}= 82%	Substrates
				Q^{2}_{inh}= 72%	Inhibitors
				$Q^{3}_{Non-int}$=89%	Non-interaction
Österberg and Norinder [144]	PLS[a]	MolSurf descriptors	N = 22	$Q^{4}_{overall}$= 51%	Substrates
Bakken and Jurs [145]	GA[b]/LDA[c]	Topological, geometric, polar surface,	N = 609	$Q_{overall}$=83-92%	Inhibitors
Dearden *et al.* [146]	MLR[d]	ClogP, topological, H-bond	N = 22	$Q_{overall}$=74.2%	Substrates
Österberg and Norinder [147]	PLS	LogP, MW, molar refraction, molar volume, parachor, polarizability, density	N = 29	$Q_{overall}$=62.7-70.6%	Substrates
Ekins *et al.* [148]	*pharmacophore model*	Catalyst version 4.5 index	N^{ii}_{Set1}= 27	Q^{5}_{set1}= 77%	Inhibitors
			N_{Set2}=21	Q_{set2}= 88%	
			N_{Set3}=17	Q_{set3}= 86%	
			N_{Set4}=18	Q_{set4}= 76%	
Onishi *et al.* [149]	MLR	Fragment codes	N = 41	$Q_{overall}$= 95.2%-95.4%	Substrates
Wang *et al.* [150]	MLR	2D descriptors 3D descriptors	N_{Set1}=22 N_{Set2}=16 N_{Set3}=11	Q_{set1}=86.3% Q_{set2}=74.2% Q_{set3}=73.6%	Inhibitors Substrates
Xue *et al.*[82]	SVM[e]	Simple molecular properties, molecular connectivity and shape, electrotopological state, quantum chemical properties, geometrical	N = 201	Q_{Subs}= 81.2%	Substrates
				$Q^{6}_{Non-Subs}$=79.2%	Non-substrates
Gombar *et al.* [81]	LDA	MW, H-bond, lipophilicity, electrotopological, molecular bulk, molar refraction	N= 153 (95/58)	$Q_{overall}$=86.2%	Substrates
Wang *et al.* [151]	Kohonen self-organizing maps (SOM)	Molecular connectivity, electrotopological state, H-bond	N= 206	Q_{inh}=80.8%	Inhibitors
				Q_{Subs}=83.3%	Substrates
Cianchetta *et al.* [152]	PLS/ pharmacophore model	Physicochemical-based descriptors	N= 129	$Q_{overall}$=74%	Substrates
Lima *et al.* [85]	*k*-NN[f], decision tree, binary QSAR, and SVM	Molecular connectivity, atom pair, VolSurf descriptors, molecular operation environment	N= 195	Q^{7}_{tr}=88-94% Q^{8}_{test}=81%	Substrates
Authors	Methods	Descriptors	Number of compounds	Overall accuracy	Interaction

(Table 5). Contd…..

Authors	Methods	Descriptors	Number of Compounds	Overall Accuracy	Interaction
Crivori *et al*. [153]	PLSD[g]	VolSurf descriptors	N^{iii}_{Tr}= 53	Q_{tr}= 88.7%	Inhibitors Substrates
			N^{iv}_{test}=272	Q_{test}= 72%	
Cabrera *et al*.[83]	LDA	TOPS-MODE topological descriptors	N= 163	$Q_{overall}$=80.9%	Substrates
			N^{v}_{ext}= 40	Q^{9}_{ext}=77.5%	
Huang *et al*. [84]	MLR SVM	Dragon descriptors (0D-3D)	N= 203	$Q_{overall}$=95.5%	Substrates
			N_{ext}= 40	Q_{ext}=93%	
Sheu *et al*. [154]	MLR	Physicochemical-based descriptors	N= 22	Correlation: R^2=0.48	Enhanced
				R^2=0.68	Low inhibitory
				R^2=0.59	High inhibitory
Estrada *et al*.[155]	-	TOPS-MODE topological descriptors	- N_{ext}= 177	Statistically significant Q_{ext}=81.8%	Substrates

[a]Partial least square; [b]genetic algorithm; [c]linear discriminant analysis; [d]multiple linear regression; [e]support vector machine; [f]k-nearest neighbor; [g]partial least squares discriminant analysis; [i]overall training and test set to model building; [ii]different sets of compounds belonging to the same structural family or metabolism action; [iii]number of compounds used for training set; [vi]number of compounds used for test set; [v]number of compounds used for external test set; [1-3,6]accuracy of classification model for substrates, inhibitor and non-interaction compounds or non-substrates; [4,7-9]overall accuracy of correlation or classification of training set and/or test set and external set; [5]accuracy of local model for a specified group of compounds.

approaches or docking methods might give the best overall models assisting in the design of new drugs [73, 74].

Finally, the prediction issue considers that the accuracy of model prediction must be experimentally validated. Therefore, suitable methods and species should be selected with precautions. Models based on data from a specific species cannot be evaluated through experiments on other species due to the lack of correlation. As a primary cause of the data deviations, Iwatsubo *et al.* suggested that interindividual variability of *in vitro* values should be corrected by using a scaling factor estimated from the metabolism of typical substrates [75].

3.2. P-Glycoprotein (P-gp)

P-gp is an ATP-dependent efflux transporter that affects the absorption, distribution and excretion of some clinically important drugs [76]. Because of the broad impact of this drug efflux transporter on *in vivo* disposition and pharmacokinetics, the identification of compounds that are P-gp substrates can help with the selection and the optimization of new drug candidates in the early stage of drug development [77]. Several "*in silico"* approaches have been developed to classify compounds as P-gp substrates [78-85]. However, no single QSPR or pharmacophore model can describe the spatial arrangement of structural features responsible for substrate and inhibitor affinity [78, 86].

As can be seen in Table **5**, most of the current works are based on relative small databases and their performances still need to be improved. Similar to QSPR models of CYP3A4, the P-gp predictions have been developed without considering their relationship with oral bioavailability, which limit their applicability during the drug discovery process. A good strategy to get insight of the role of P-gp on drug bioavailability could be to develop *in silico* models of P-gp substrates combined with computational models of solubility and permeability.

On the other hand, rule-based approaches should receive significant attention because a filter prior to screening a large dataset is needed. Considering the active efflux alone, Didziapetris *et al*. used a set of 220 compounds and proposed the "rule of four" [87], which states that compounds are likely to be efflux substrates if they have a hydrogen bond acceptor count (sum of N and O atoms) ≥ 8, MW > 400 and an acid group with pKa > 4. Conversely, compounds are likely to be non-substrates if they have an acceptor count ≤ 4, MW < 400 and a base witpKa < 8. Gleeson reported that neutral or basic molecules showing a MW > 400 and a logP > 4 are more likely to be transported by P-gp than acidic or zwitterionic compounds [88].

4. ORAL BIOAVAILABILITY PREDICTION

During the last ten years, different computational models to predict oral bioavailability have been reported (See Table **6**). These predictions are much more difficult than those for intestinal absorption or permeability because oral bioavailability primarily depends on a superposition of intestinal absorption and first-pass metabolism. On the other hand, the previous discussed QSPR models for individual properties related to oral bioavailability such as solubility, intestinal permeability, human intestinal absorption and substrates of P-gp and/or CYP3A4 are not able to explain the multi-factorial nature of oral bioavailability. At this point an important question arise: *Is it better to use specific or integral QSPR methodologies for the prediction of oral*

Table 6. Different *in Silico* Models to Predict Oral Bioavailability

Reference	Statistical Method	Descriptors	Database	Performance of the Best Model
Andrews [94]	MLR[a]	85 fragment	N^{i}= 591	$Q^{1}_{training}$= 0.71 Q^{3}_{CV} (LOO)= 0.63 Q^{4}_{CV} (LGO)= 0.58 (80/20)
Yoshida [156]	ORMUCS[b] method	Log*D*6.5, (log*D*6.5)^2, Δlog*D*, and 15 fragment	N^{iii}_{tr}= 232 N^{iv}_{test}= 40	$Q_{training}$= 71% Q_{CV} (LOO)= 67% Q_{test}= 60%
Turner [157]	MLR	Physicochemical properties, topological, constitutional, geometrical and quantum chemical descriptors	N_{tr}= 159 N_{test}= 10	$Q_{training}$= 0.35 Q_{CV} (LOO)= 0.25 Q^{2}_{test}= 0.72
Pintore [91]	Adaptive fuzzy partitioning	Ten molecular descriptors	N^{ii}_{Data1}= 272 N_{Data2}= 432	$Q_{traning1}$= 82% $Q_{validation1}$ = 75% Q_{test1}= 40% $Q_{traning2}$= 70% $Q_{validation2}$ = 68% Q_{test2}= 64%
Turner [92]	ANN[c]	Physicochemical properties, topological, constitutional, geometrical and quantum chemical descriptors	N= 167	$Q_{training}$=0.74 $Q_{validation}$ = 0.90 Q_{test}= 0.68
Wang [58]	GA[d]-QSPR	Multiple molecular descriptors	N= 577	$Q_{training}$= 0.55 Q_{CV} (LGO)= 0.42 (90/10)
Hou [55]	MLR	Molecular properties	N=678	No good rules for predicting oral bioavailability
Moda *et al*. [93]	HQSPR (hologram)	Multiple molecular descriptors	N_{tr}=250 N_{test}=52	$Q_{training}$= 0.93 Q_{CV} = 0.70 Q_{test}= 0.85
Chang-Ying *et al*. [90]	GA-CG[e]-SVM[f]	Multiple molecular descriptors	N_{tr}=690 N_{test}=76	$Q_{training}$= 0.80 Q_{test}= 0.86

[a]Multiple linear regression; [b]ordered multicategorical classification method using the simplex technique; [c]artificial neural network; [d]genetic algorithm; [e]conjugate gradient; [f]support vector machine; i,ii [i]overall of training and test set to model building; [iii]number of compounds used for training set; [vi]number of compounds used for test set; [1,2]overall accuracy of correlation or classification of training set and/or test set; [3,4]accuracy of leave-one-out cross-validation or leave-group-out cross validation.

bioavailability? Finding a clear answer to this question is still a challenge!

From these models one can see the difficulty encountered when establishing a relationship between oral bioavailability and simple molecular descriptors. Some authors have tried to improve the bioavailability prediction including experimental absorption data such as HIA as a descriptor of the bioavailability model [89]. This strategy has been carried out to skip all the initial factors related to drug dissolution and absorption processes. However, the same authors have pointed out that experimental HIA values are still difficult to obtain, being available for only a limited set of compounds. In this situation, the use of HIA values predicted by a computational model is mandatory. Other authors suggest using nonlinear learning method to overcome the problem of bioavailability prediction [90].

Although relatively large datasets have been used to develop *in silico* models of human oral bioavailability, the results achieved are not really good. In some models the predictive capacity is questionable due to the low correlation of the cross-validation and in others the kind of descriptor limits the interpretation of bioavailability.

5. PERSPECTIVES AND CONCLUSIONS

The prediction of ADME properties has always been challenge for drug design researchers. During the last decade many attempts have been made to explore the quantitative and/or qualitative relationship of structure-oral bioavailability and different predictive results have been

achieved [53-55, 58, 89, 91-93]. Nowadays there is an ongoing discussion regarding the generation of new *in silico* models to predict oral human bioavailability or to refine the reported models including information related to absorption and metabolic processes.

Several factors are involved in the poor computational prediction of oral bioavailability and most of them are related to the experimental dataset, the molecular descriptors or the statistical methodology used to obtain the models. Concerning datasets, the main problem is their small size and the lack of extensive and reliable experimental data. In the first case, the models have a limited predictive capacity due to the poor structural variability of compounds in the training sets while the second one points out the need to develop reliable and reproducible absorption models. Some models have been developed using extensive and reliable datasets belonging to big pharmaceutical companies [53, 94]. Unfortunately raw data was not available to the scientific community, which limited the possibility to obtain better predictive absorption models. Recently, two extensive databases were published (http://modem.ucsd.edu/adme/ and http://miro.ifsc.usp.br/pkdb/), bringing some opportunities to study the different absorption properties and to increase, in the future, the quality of prediction models.

Different molecular descriptors have been used to explain the relationship between chemical structure and oral bioavailability but none characterize effectively the first-pass metabolism [95]. Concerning this, some authors have suggested the introduction of new rules or molecular descriptors to improve the *in silico* bioavailability models [48, 55]. In this case, the use of topological descriptors with fragment/bond information like TOPS-MODE [96] could be interesting in the prediction of oral bioavailability.

In our opinion, the use of a dataset, carefully selected, with information about the most important processes involved in bioavailability in order to develop different models (consensus model) to predict oral bioavailability can be a valuable strategy. For example, select compounds with F and HIA values reported, and classify the dataset considering from these compounds those that are or not substrates of P-gp (efflux process) and/or substrates of CYP3A4 (intestinal and hepatic metabolism).

In this sense, we want to highlight several important aspects to be considered in the future development of bioavailability computational models:

1. The quality and quantity of available databases built from reliable experimental assays have to be improved.
2. The inter- and intra-laboratory variability has to be taken into account with an appropriate judgment.
3. The active transport should be considered.
4. The combination of two or more models for the same property, based on different principles, could provide a high confidence in the outcome and assist decision-making. The new era of QSPR should pave the way for a multi-objective optimization with novel technologies applied in oral bioavailability problem.
5. *In silico* results must be validated at least by *in vitro* experimental assays, and if possible, also by *in situ* and *in vivo* assays.

In this review, we have executed an intensive analysis of progress of computational models in the prediction of oral bioavailability and related properties. The main factors that affect the absorption process have also been discussed to reveal that oral absorption is a multi-factorial process and its prediction can be divided into three independent tasks: (**1**) prediction of solubility related dose of drug; (**2**) estimation of intestinal permeability and (**3**) development of an adequate tool to identify substrate structure of efflux and biotransformation, especially by CYP450, in human. Although significant efforts have been made in the use of *in vivo*, *in situ* and *in vitro* assays to predict absorption and bioavailability, *in silico* methods to predict these characteristics are expected to decrease time and money consuming experiments and to increase the power of screen compounds prior to synthesis.

Finally, it is important to mention that *in silico* prediction of oral bioavailability is really complex and although novel approaches have been carried out, the development of reliable and predictive *in silico* tools is still a challenge for the computational design experts.

ACKNOWLEDGEMENTS

The authors acknowledge financial support of Agencia Española de Cooperación Iberoamericana para el Desarrollo (AECID) to the project D/024153/09: *Montaje de un laboratorio de Química Computacional, con fines académicos y científicos, para el diseño de potenciales candidatos a fármacos, en enfermedades de alto impacto social* and the project DCI-ALA/ALFA III (2010) 29: *Red para el desarrollo de metodologías biofarmacéuticas racionales que incrementen la competencia y el impacto social de las Industrias Farmacéuticas Locales. (Red-Biofarma)* M.A.C.P also thanks to the program: *Estancias de movilidad de profesores e investigadores extranjeros en centros españoles (SAB2009-0106),* developed at Miguel Hernández University and to Christiane Mathes for reviewing the English text.

REFERENCES

[1] Prentis, R.A.; Lis, Y.; Walker, S.R., Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964-1985). *Br J Clin Pharmacol,* **1988**, *25*(3), 387-396.
[2] Kennedy, T., Managing the drug discovery/ development interface. *Drug Discov Today,* **1997**, *2*, 436-444.
[3] Kubinyi, H., Drug research: myths, hype and reality. *Nat Rev Drug Discov,* **2003**, *2*(8), 665-668.
[4] Han van de Waterbeemd; Testa, B. *Drug Bioavailability. Estimation of Solubility, Permeability, Absorption and Bioavailability*. Second ed. WILEY-VCH: Weinheim, **2008**.
[5] Hou, T.; Li, Y.; Zhang, W.; Wang, J., Recent developments of in silico predictions of intestinal absorption and oral bioavailability. *Comb Chem High Throughput Screen,* **2009**, *12*(5), 497-506.
[6] Manitpisitkul, P.; White, R.E., Whatever happened to cassette-dosing pharmacokinetics? *Drug Discov Today,* **2004**, *9*(15), 652-658.
[7] Artursson, P.; Palm, K.; Luthman, K., Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv Drug Deliv Rev,* **2001**, *46*(1-3), 27-43.
[8] Li, A.P., Screening for human ADME/Tox drug properties in drug discovery. *Drug Discov Today,* **2001**, *6*(7), 357-366.

[9] Nordqvist, A.; Nilsson, J.; Lindmark, T.; Eriksson, A.; Garberg, P.; Kihlen, M., A General Model for Prediction of Caco-2 Cell Permeability. *QSAR Comb Sci,* **2004**, *23*, 303-310.

[10] Roberts, S.A., High-throughput screening approaches for investigating drug metabolism and pharmacokinetics. *Xenobiotica,* **2001**, *31*(8-9), 557-589.

[11] Avdeef, A. *Absorption and Drug Development: Solubility, Permeability, and Charge State*. John Wiley & Sons, Inc, **2003**.

[12] Burton, P.S.; Goodwin, J.T.; Vidmar, T.J.; Amore, B.M., Predicting drug absorption: how nature made it a difficult problem. *J Pharmacol Exp Ther,* **2002**, *303*(3), 889-895.

[13] http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=320. (17 February),

[14] http://www.emea.europa.eu/pdfs/human/ewp/140198en.pdf. (10 February),

[15] Navarro-Fontestad, C.; Gonzalez-Alvarez, I.; Fernandez-Teruel, C.; Bermejo, M.; Casabo, V.G., A new mathematical approach for the estimation of the AUC and its variability under different experimental designs in preclinical studies. *Pharm Stat,* **2011**, in press.

[16] Lin, J.H.; Chiba, H.; Baillie, T.A., Is the role of the small intestine in first-pass metabolism overemphasized? *Pharmacol Rev.,* **1999**, *51*(2), 135-158.

[17] Chaturvedi, P.R.; Decker, C.J.; Odinecs, A., Prediction of pharmacokinetic properties using experimental approaches during early drug discovery. *Curr Opin Chem Biol,* **2001**, *5*(4), 452-463.

[18] Wacher, V.J.; Salphati, L.; Benet, L.Z., Active secretion and enterocytic drug metabolism barriers to drug absorption. *Adv Drug Deliv Rev,* **2001**, *46*(1-3), 89-102.

[19] Kharasch, E.D.; Walker, A.; Hoffer, C.; Sheffels, P., Evaluation of First-Pass Cytochrome P4503A (CYP3A) and P-Glycoprotein Activities Using Alfentanil and Fexofenadine in Combination. *J Clin Pharmacol,* **2005**, *45*(1), 79-88.

[20] Benet, L.Z.; Cummins, C.L., The drug efflux-metabolism alliance: biochemical aspects. *Adv Drug Deliv Rev,* **2001**, *50 Suppl 1*, S3-11.

[21] Amidon, G.L.; Lennernas, H.; Shah, V.P.; Crison, J.R., A theoretical basis for a biopharmaceutic drug classification: the correlation of *in vitro* drug product dissolution and *in vivo* bioavailability. *Pharm Res,* **1995**, *12*(3), 413-420.

[22] FDA Guidance for Industry: Federal Drug and Food Administration: Rockville, MD, **2002**.

[23] Kwan, K.C., Oral bioavailability and first-pass effects. *Drug Metab Dispos,* **1997**, *25*(12), 1329-1336.

[24] Oh, D.M.; Curl, R.L.; Amidon, G.L., Estimating the fraction dose absorbed from suspensions of poorly soluble compounds in humans: a mathematical model. *Pharm Res,* **1993**, *10*(2), 264-270.

[25] Norinder, U.; Bergstrom, C.A., Prediction of ADMET Properties. *ChemMedChem,* **2006**, *1*(9), 920-937.

[26] Johnson, S.R.; Zheng, W., Recent progress in the computational prediction of aqueous solubility and absorption. *Aaps J,* **2006**, *8*(1), E27-40.

[27] Bergstrom, C.A.; Wassvik, C.M.; Norinder, U.; Luthman, K.; Artursson, P., Global and Local Computational Models for Aqueous Solubility Prediction of Drug-Like Molecules. *J Chem Inf Comput Sci,* **2004**, *44*(4), 1477-1488.

[28] Ghasemi, J.; Saaidpour, S., QSPR prediction of aqueous solubility of drug-like organic compounds. *Chem Pharm Bull (Tokyo),* **2007**, *55*(4), 669-674.

[29] Huuskonen, J.; Livingstone, D.J.; Manallack, D.T., Prediction of drug solubility from molecular structure using a drug-like training set. *SAR QSAR Environ Res,* **2008**, *19*(3-4), 191-212.

[30] Delaney, J.S., ESOL: estimating aqueous solubility directly from molecular structure. *J Chem Inf Comput Sci,* **2004**, *44*(3), 1000-1005.

[31] Votano, J.R.; Parham, M.; Hall, L.H.; Kier, L.B.; Hall, L.M., Prediction of aqueous solubility based on large datasets using several QSPR models utilizing topological structure representation. *Chem Biodivers,* **2004**, *1*(11), 1829-1841.

[32] Wang, J.; Hou, T.; Xu, X., Aqueous solubility prediction based on weighted atom type counts and solvent accessible surface areas. *J Chem Inf Model,* **2009**, *49*(3), 571-581.

[33] Johnson, S.R.; Chen, X.Q.; Murphy, D.; Gudmundsson, O., A computational model for the prediction of aqueous solubility that includes crystal packing, intrinsic solubility, and ionization effects. *Mol Pharm,* **2007**, *4*(4), 513-523.

[34] Hidalgo, I.J.; Raub, T.J.; Borchardt, R.T., Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability. *Gastroenterology,* **1989**, *96*(3), 736-749.

[35] Yamashita, S.; Tanaka, Y.; Endoh, Y.; Taki, Y.; Sakane, T.; Nadai, T.; Sezaki, H., Analysis of drug permeation across Caco-2 monolayer: implication for predicting *in vivo* drug absorption. *Pharm Res,* **1997**, *14*(4), 486-491.

[36] Yee, S., *In vitro* permeability across Caco-2 cells (colonic) can predict *in vivo* (small intestinal) absorption in man--fact or myth. *Pharm Res,* **1997**, *14*(6), 763-766.

[37] Wahlang, B.; Patil, S.R.; Pawar, Y.B.; Bansal, A.K., The Caco-2 Cell Model: A useful Tool in Drug Dicovery and Development. *CRIPS,* **2009**, *10*(2), 29-34.

[38] Delie, F.; Rubas, W., A human colonic cell line sharing similarities with enterocytes as a model to examine oral absorption: advantages and limitations of the Caco-2 model. *Crit Rev Ther Drug Carrier Syst,* **1997**, *14*(3), 221-286.

[39] Artursson, P.; Palm, K.; Luthman, K., Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv Drug Deliv Rev,* **1996**, *22*, 67-84.

[40] Yamashita, F.; Fujiwara, S.; Hashida, M., The "latent membrane permeability" concept: QSPR analysis of inter/intralaboratory variable Caco-2 permeability. *J Chem Inf Comput Sci,* **2002**, *42*(2), 408-413.

[41] Kerns, E.H.; Di, L. *Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization*. Elsevier's Science & Technology Rights Department in Oxford: 84 Theobald's Road, London WC1X 8RR, UK, **2008**.

[42] Pham-The, H.; González-Álvarez, I.; Bermejo, M.; Mangas, V.; Centelles, I.; Garrigues, T.M.; Cabrera-Pérez, M.A., In-silico prediction of Caco-2 cell permeability by a classification QSAR approach. *Mol Inf.* , **2011**, *30*(4), 376-385.

[43] Wu, C.Y.; Benet, L.Z., Predicting drug disposition via application of BCS: transport/absorption/ elimination interplay and development of a biopharmaceutics drug disposition classification system. *Pharm Res,* **2005**, *22*(1), 11-23.

[44] Bergstrom, C.A.; Strafford, M.; Lazorova, L.; Avdeef, A.; Luthman, K.; Artursson, P., Absorption classification of oral drugs based on molecular surface properties. *J Med Chem,* **2003**, *46*(4), 558-570.

[45] Bergstrom, C.A., *In silico* predictions of drug solubility and permeability: Two rate-limiting barriers to oral drug absorption. *Basic Clin Pharmacol Toxicol,* **2005**, *96*, 156-161.

[46] Khandelwal, A.; Bahadduri, P.M.; Chang, C.; Polli, J.E.; Swaan, P.W.; Ekins, S., Computational models to assign biopharmaceutics drug disposition classification from molecular structure. *Pharm Res,* **2007**, *24*(12), 2249-2262.

[47] Kasim, N.A.; Whitehouse, M.; Ramachandran, C.; Bermejo Sanz, M.; Lennernas, H.; Hussain, A.S.; Junginger, H.E.; Stavchansky, S.A.; Midha, K.K.; Shah, V.P.; Amidon, G., Molecular Properties of WHO Essential Drugs and Provisional Biopharmaceutical Classification. *Mol Pharm,* **2004**, *1*(1), 85-96.

[48] Hou, T.; Wang, J.; Zhang, W.; Xu, X., ADME evaluation in drug discovery. 7. Prediction of oral absorption by correlation and classification. *J Chem Inf Model,* **2007**, *47*(1), 208-218.

[49] Sinko, P.J., Drug selection in early drug development: screening for acceptable pharmacokinetic properties using combined *in vitro* and computational approaches. *Curr Opin Drug Discov Devel,* **1999**, *2*(1), 42-48.

[50] Tsaioun, K.; Bottlaender, M.; Mabondzo, A., ADDME - Avoiding Drug Development Mistakes Early: central nervous system drug discovery perspective. *BMC Neurol,* **2009**, *9 Suppl 1*, S1.

[51] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J., Experimental and computational approaches to estimate solubility and permeability in the drug discovery and development settings. *Adv Drug Deliv Rev,* **1997**, *23*, 3-25.

[52] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H., A rule of three for fragment – based lead discovery? . *Drug Discov. Today,* **2003**, *8*, 876 – 877.

[53] Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D., Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem,* **2002**, *45*(12), 2615-2623.

[54] Lu, J.J.; Crimin, K.; Goodwin, J.T.; Crivori, P.; Orrenius, C.; Xing, L.; Tandler, P.J.; Vidmar, T.J.; Amore, B.M.; Wilson, A.G.; Stouten, P.F.; Burton, P.F., Influence of Molecular Flexibility and Polar Surface Area Metrics on Oral Bioavailability in the Rat. *J Med Chem,* **2004**, *47*, 6104-6107.

[55] Hou, T.; Wang, J.; Zhang, W.; Xu, X., ADME evaluation in drug discovery. 6. Can oral bioavailability in humans be effectively predicted by simple molecular property-based rules? *J Chem Inf Model,* **2007**, *47*(2), 460-463.

[56] Giménez, B.G.; Santos, M.S.; Ferrarini, M.; Fernandes, J.P.S., Evaluation of blockbuster drugs under the Rule-of-five. *Pharmazie,* **2010**, *65*, 148–152.

[57] Overington, J.P.; Al-Lazikani, B.; Hopkins, A.L., How many drug targets are there? *Nat Rev Drug Discov. ,* **2006**, *5*(12), 993-996.

[58] Wang, J.; Krudy, G.; Xie, X.Q.; Wu, C.; Holland, G., Genetic algorithm-optimized QSPR models for bioavailability, protein binding, and urinary excretion. *J Chem Inf Model,* **2006**, *46*(6), 2674-2683.

[59] Lown, K.S.; Mayo, R.R.; Leichtman, A.B.; Hsiao, H.L.; Turgeon, D.K.; Schmiedlin, R.P.; Brown, M.B.; Guo, W.; Rossi, S.J.; Benet, L.Z., Role of intestinal P-glycoprotein (mdr 1) in interpatient variation in the oral bioavailability of cyclosporine. *Clin. Pharmacol. Ther. ,* **1997**, *62*, 248–260.

[60] Lown, K.S.; Kolars, J.C.; Thummel, K.E.; Barnett, J.L.; Kunze, K.L.; Wrighton, S.A.; Watkins, P.B., Interpatient heterogeneity in expression of CYP3A4 and CYP3A5 in small bowel. Lack of prediction by the erythromycin breath test. *Drug Metabol. Dispos. ,* **1994**, *22*, 947–955.

[61] Guengerich, F.P. *Cytochrome P450 enzymes, in Cytochrome P450: Structure, Mechanism and Biochemistry*. Plenum Press: New York, **1995**.

[62] Wang, E.; Lew, K.; Barecki, M.; Casciano, C.N.; Clement, R.P.; Johnson, W.W., Quantitative distinctions of active site molecular recognition by P-glycoprotein and cytochrome P450 3A4. *Chem Res Toxicol,* **2001**, *14*(12), 1596-1603.

[63] Lill, M.A.; Dobler, M.; Vedani, A., Prediction of small-molecule binding to cytochrome P450 3A4: flexible docking combined with multidimensional QSAR. *ChemMedChem,* **2006**, *1*(1), 73-81.

[64] Ekins, S.; Bravi, G.; Wikel, J.H.; Wrighton, S.A., Three-Dimensional-Quantitative Structure Activity Relationship Analysis of Cytochrome P-450 3A4 Substrates. *J. Pharmacol. Exp. Ther.,* **1999**, *291*(1), 424-433.

[65] Ekins, S.; Bravi, G.; Binkley, S.; Gillespie, J.S.; Ring, B.J.; Wikel, J.H.; Wrighton, S.A., Three- and Four-Dimensional Quantitative Structure Activity Relationship Analyses of Cytochrome P-450 3A4 Inhibitors. *J Pharmacol Exp Ther,* **1999**, *290*, 429-438.

[66] Ekins, S.; de Groot, M.J.; Jones, J.P., Pharmacophore and three-dimensional quantitative structure activity relationship methods for modeling cytochrome p450 active sites. *Drug Metab Dispos,* **2001**, *29*(7), 936-944.

[67] Lewis, D.F.; Lake, B.G.; Dickins, M., Quantitative structure-activity relationships (QSARs) in CYP3A4 inhibitors: the importance of lipophilic character and hydrogen bonding. *J Enzyme Inhib Med Chem.,* **2006**, *21*(2), 127-132.

[68] Sheridan, R.P.; Korzekwa, K.R.; Torres, R.A.; Walker, M.J., Empirical regioselectivity models for human cytochromes P450 3A4, 2D6, and 2C9. *J Med Chem,* **2007**, *14*(50), 3173-3184.

[69] Singh, S.B.; Shen, L.Q.; Walker, M.J.; Sheridan, R.P., A Model for Predicting Likely Sites of CYP3A4-mediated Metabolism on Drug-like Molecules. *J. Med. Chem.,* **2003,** *46*(8),1330-1336.

[70] Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianello, R., MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J Med Chem,* **2005**, *48*(22), 6970-6979.

[71] Zheng, M.; Luo, X.; Shen, Q.; Wang, Y.; Du, Y.; Zhu, W.; Jiang, H., Site of metabolism prediction for six biotransformations mediated by cytochromes P450. *Bioinformatics,* **2009**, *25*(10), 1251-1258.

[72] Li, H.; Sun, J.; Fan, X.; Sui, X.; Zhang, L.; Wang, Y.; He, Z., Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction. *J Comput Aided Mol Des,* **2008**, *22*(11), 843-855.

[73] Feher, M.; Ewing, T., Global or Local QSAR: Is There a Way Out? *QSAR Comb. Sci.,* **2009**, *28*(8), 850-855.

[74] Li, H.; Sun, J.; Fan, X.; Sui, X.; Zhang, L.; Wang, Y.; He, Z., Considerations and recent advances in QSAR models for cytochrome P450-mediated drug metabolism prediction. *J Comput Aided Mol Des,* **2008**, *22*(11), 843-855.

[75] Iwatsubo, T.; Hirota, N.; Ooie, T.; Suzuki, H.; Shimada, N.; Chiba, K.; Ishizaki, T.; Green, C.E.; Tyson, C.A.; Sugiyama, Y., Prediction of *in vivo* drug metabolism in the human liver from *in vitro* metabolism data. *Pharmacol Ther,* **1997**, *73*(2), 147-171.

[76] Schinkel, A.H., P-Glycoprotein, a gatekeeper in the blood–brain barrier. *Adv. Drug Deliver. Rev.,* **1999**, *36*, 179-194.

[77] Polli, J.W.; Wring, S.A.; Humphreys, J.E.; Huang, L.; Morgan, J.B.; Webster, L.O.; Serabjit-Singh, C.S., Rational use of *in vitro* P-glycoprotein assays in drug discovery. *J Pharmacol Exp Ther,* **2001**, *299*(2), 620-628.

[78] Stouch, T.R.; Gudmundsson, O., Progress in understanding the structure-activity relationships of P-glycoprotein. *Adv Drug Deliv Rev,* **2002**, *54*(3), 315-328.

[79] Ekins, S.; Waller, C.L.; Swaan, P.W.; Cruciani, G.; Wrighton, S.A.; Wikel, J.H., Progress in predicting human ADME parameters in silico. *J Pharmacol Toxicol Methods,* **2000**, *44*(1), 251-272.

[80] Zhang, E.Y.; Phelps, M.A.; Cheng, C.; Ekins, S.; Swaan, P.W., Modeling of active transport systems. *Adv Drug Deliv Rev,* **2002**, *54*(3), 329-354.

[81] Gombar, V.K.; Polli, J.W.; Humphreys, J.E.; Wring, S.A.; Serabjit-Singh, C.S., Predicting P-glycoprotein substrates by a quantitative structure-activity relationship model. *J Pharm Sci,* **2004**, *93*(4), 957-968.

[82] Xue, Y.; Yap, C.W.; Sun, L.Z.; Cao, Z.W.; Wang, J.F.; Chen, Y.Z., Prediction of P-Glycoprotein Substrates by a Support Vector Machine Approach. *J Chem Inf Comput Sci,* **2004**, *44*(4), 1497-1505.

[83] Cabrera, M.A.; Gonzalez, I.; Fernandez, C.; Navarro, C.; Bermejo, M., A topological substructural approach for the prediction of P-glycoprotein substrates. *J Pharm Sci,* **2006**, *95*(3), 589-606.

[84] Huang, J.; Ma, G.; Muhammad, I.; Cheng, Y., Identifying P-glycoprotein substrates using a support vector machine optimized by a particle swarm. *J Chem Inf Model,* **2007**, *47*(4), 1638-1647.

[85] Cerqueira-Lima, P.; Golbraikh, A.; Oloff, S.; Xiao, Y.; Tropsha, A., Combinatorial QSAR Modeling of P-Glycoprotein Substrates. *J Chem Inf Model,* **2006**, *46*(3), 1245-1254.

[86] Ekins, S.; Swaan, P., Development of computational models for enzymes, transporters, channels and receptors relevant to ADME/TOX. *Rev Comp Chem,* **2004**, *20*, 333-415.

[87] Didziapetris, R.; Japertas, P.; Avdeef, A.; Petrauskas, A., Classification analysis of P-glycoprotein substrate specificity. *J Drug Target,* **2003**, *11*(7), 391-406.

[88] Gleeson, M.P., Generation of a set of simple, interpretable ADMET rules of thumb. *J Med Chem,* **2008**, *51*(4), 817-834.

[89] Wang, Z.; Yan, A.; Yuan, Q.; Gasteiger, J., Explorations into modeling human oral bioavailability. *Eur J Med Chem,* **2008**, *43*(11), 2442-2452.

[90] Ma, C.Y.; Yang, S.Y.; Zhang, H.; Xiang, M.L.; Huang, Q.; Wei, Y.Q., Prediction models of human plasma protein binding rate and oral bioavailability derived by using GA–CG–SVM method. *J Pharm Biomed Anal,* **2008**, *47*, 677-682.

[91] Pintore, M.; van de Waterbeemd, H.; Piclin, N.; Chretien, J.R., Prediction of oral bioavailability by adaptive fuzzy partitioning. *Eur J Med Chem,* **2003**, *38*(4), 427-431.

[92] Turner, J.V.; Maddalena, D.J.; Agatonovic-Kustrin, S., Bioavailability prediction based on molecular structure for a diverse series of drugs. *Pharm Res,* **2004**, *21*(1), 68-82.

[93] Moda, T.L.; Montanari, C.A.; Andricopulo, A.D., Hologram QSAR model for the prediction of human oral bioavailability. *Bioorg Med Chem,* **2007**, *15*(24), 7738-7745.

[94] Andrews, C.W.; Bennett, L.; Yu, L.X., Predicting human oral bioavailability of a compound: development of a novel quantitative structure-bioavailability relationship. *Pharm Res,* **2000**, *17*(6), 639-644.

[95] Hou, T.; Wang, J., Structure-ADME relationship: still a long way to go? *Expert Opin Drug Metab Toxicol,* **2008**, *4*(6), 759-770.

[96] Gutierrez, Y.; Estrada, E., MODESLAB 1.5 (Molecular DEScriptors LABoratory) for Windows. **2004**.

[97] Huuskonen, J., Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology. *J Chem Inf Comput Sci,* **2000**, *40*(3), 773-777.

[98] McElroy, N.R.; Jurs, P.C., Prediction of aqueous solubility of heteroatom-containing organic compounds from molecular structure. *J Chem Inf Comput Sci,* **2001**, *41*(5), 1237-1247.

[99] Tetko, I.V.; Tanchuk, V.Y.; Kasheva, T.N.; Villa, A.E., Estimation of aqueous solubility of chemical compounds using E-state indices. *J Chem Inf Comput Sci,* **2001**, *41*(6), 1488-1493.

[100] Liu, R.; So, S.S., Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. 1. Aqueous solubility. *J Chem Inf Comput Sci,* **2001**, *41*(6), 1633-1639.

[101] Livingstone, D.J.; Ford, M.G.; Huuskonen, J.J.; Salt, D.W., Simultaneous prediction of aqueous solubility and octanol/water partition coefficient based on descriptors derived from molecular structure. *J Comput Aided Mol Des,* **2001**, *15*(8), 741-752.

[102] Cheng, A.; Diller, D.J.; Dixon, S.L.; Egan, W.J.; Lauri, G.; Merz, K.M., Jr., Computation of the physio-chemical properties and data mining of large molecular collections. *J Comput Chem,* **2002**, *23*(1), 172-183.

[103] Gao, H.; Shanmugasundaram, V.; Lee, P., Estimation of aqueous solubility of organic compounds with QSPR approach. *Pharm Res,* **2002**, *19*(4), 497-503.

[104] Engkvist, O.; Wrede, P., High-throughput, in silico prediction of aqueous solubility based on one- and two-dimensional descriptors. *J Chem Inf Comput Sci,* **2002**, *42*(5), 1247-1249.

[105] Yan, A.; Gasteiger, J., Prediction of aqueous solubility of organic compounds by topological descriptors *QSAR Comb. Sci.,* **2003**, *22*, 821-829.

[106] Yan, A.; Gasteiger, J., Prediction of aqueous solubility of organic compounds based on a 3D structure representation. *J Chem Inf Comput Sci,* **2003**, *43*(2), 429-434.

[107] Cheng, A.; Merz, K.M., Jr., Prediction of aqueous solubility of a diverse set of compounds using quantitative structure-property relationships. *J Med Chem,* **2003**, *46*(17), 3572-3580.

[108] Butina, D.; Gola, J.M., Modeling aqueous solubility. *J Chem Inf Comput Sci,* **2003**, *43*(3), 837-841.

[109] Manallack, D.T.; Tehan, B.G.; Gancia, E.; Hudson, B.D.; Ford, M.G.; Livingstone, D.J.; Whitley, D.C.; Pitt, W.R., A consensus neural network-based technique for discriminating soluble and poorly soluble compounds. *J Chem Inf Comput Sci,* **2003**, *43*(2), 674-679.

[110] Schaper, K.J.; Kunz, B.; Raevsky, O.A., Analysis of water solubility data on the basis of HYBOT descriptors, part 2: solubility of liquid chemicals and drugs. . *QSAR Comb. Sci.,* **2003**, *22*, 943-958.

[111] Wegner, J.K.; Zell, A., Prediction of aqueous solubility and partition coefficient optimized by a genetic algorithm based descriptor selection method. *J Chem Inf Comput Sci,* **2003**, *43*(3), 1077-1084.

[112] Hou, T.J.; Xia, K.; Zhang, W.; Xu, X.J., ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach. *J Chem Inf Comput Sci,* **2004**, *44*(1), 266-275.

[113] Wang, J.; Krudy, G.; Hou, T.; Zhang, W.; Holland, G.; Xu, X., Development of reliable aqueous solubility models and their application in druglike analysis. *J Chem Inf Model,* **2007**, *47*(4), 1395-1404.

[114] Fujiwara, S.; Yamashita, F.; Hashida, M., Prediction of Caco-2 cell permeability using a combination of MO-calculation and neural network. *Int J Pharm,* **2002**, *237*(1-2), 95-105.

[115] Yamashita, F.; Wanchana, S.; Hashida, M., Quantitative structure/property relationship analysis of Caco-2 permeability using a genetic algorithm-based partial least squares method. *J Pharm Sci,* **2002**, *91*(10), 2230-2239.

[116] Hou, T.J.; Zhang, W.; Xia, K.; Qiao, X.B.; Xu, X.J., ADME Evaluation in Drug Discovery. 5. Correlation of Caco-2 Permeation with Simple Molecular Properties. *J. Chem. Inf. Comput. Sci.,* **2004**, *44*, 1585-1600.

[117] Marrero-Ponce, Y.; Cabrera, M.A.; Romero, V.; Bermejo, M.; Siverio, D.; Torrens, F., Prediction of Intestinal Epithelial Transport of Drug in (Caco-2) Cell Culture from Molecular Structure using in silico Approaches During Early Drug Discovery. *Internet Electron J Mol Des,* **2005**, *4*(2), 124-150.

[118] Refsgaard, H.H.; Jensen, B.F.; Brockhoff, P.B.; Padkjaer, S.B.; Guldbrandt, M.; Christensen, M.S., In silico prediction of membrane permeability from calculated molecular parameters. *J Med Chem,* **2005**, *48*(3), 805-811.

[119] Guangli, M.; Yiyu, C., Predicting Caco-2 permeability using support vector machine and chemistry development kit. *J Pharm Pharm Sci,* **2006**, *9*(2), 210-221.

[120] Di Fenza, A.; Alagona, G.; Ghio, C.; Leonardi, R.; Giolitti, A.; Madami, A., Caco-2 cell permeability modelling: a neural network coupled genetic algorithm approach. *J Comput Aided Mol Des,* **2007**, *21*(4), 207-221.

[121] Castillo-Garit, J.A.; Marrero-Ponce, Y.; Torrens, F.; Garcia-Domenech, R., Estimation of ADME properties in drug discovery: predicting Caco-2 cell permeability using atom-based stochastic and non-stochastic linear indices. *J Pharm Sci,* **2008**, *97*(5), 1946-1976.

[122] Paixao, P.; Gouveia, L.F.; Morais, J.A., Prediction of the *in vitro* permeability determined in Caco-2 cells by using artificial neural networks. *Eur J Pharm Sci,* **2010**, *41*(1), 107-117.

[123] Zhao, Y.H.; Le, J.; Abraham, M.H.; Hersey, A.; Eddershaw, P.J.; Luscombe, C.N.; Butina, D.; Beck, G.; Sherborne, B.; Cooper, I.; Platts, J.A., Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure-activity relationship (QSAR) with the Abraham descriptors. *J Pharm Sci,* **2001**, *90*(6), 749-784.

[124] Klopman, G.; Stefan, L.R.; Saiakhov, R.D., ADME evaluation. 2. A computer model for the prediction of intestinal absorption in humans. *Eur J Pharm Sci,* **2002**, *17*(4-5), 253-263.

[125] Cabrera, M.A.; Bermejo, M.; Ramos, L.; Grau, R.; Pérez-González, M.; González-Díaz, H., A topological sub-structural approach for predicting human intestinal absorption of drugs. *Eur J Med Chem,* **2004**, *39*(11), 905-916.

[126] Deconinck, E.; Hancock, T.; Coomans, D.; Massart, D.L.; Heyden, Y.V., Classification of drugs in absorption classes using the classification and regression trees (CART) methodology. *J Pharm Biomed Anal,* **2005**, *39*(1-2), 91-103.

[127] Liu, H.X.; Hu, R.J.; Zhang, R.S.; Yao, X.J.; Liu, M.C.; Hu, Z.D.; Fan, B.T., The prediction of human oral absorption for diffusion rate-limited drugs based on heuristic method and support vector machine. *J Comput Aided Mol Des,* **2005**, *19*(1), 33-46.

[128] Hou, T.; Wang, J.; Li, Y., ADME evaluation in drug discovery. 8. The prediction of human intestinal absorption by a support vector machine. *J Chem Inf Model,* **2007**, *47*(6), 2408-2415.

[129] Gunturi, S.B.; Narayanan, R., In Silico ADME Modeling 3: Computational Models to Predict Human Intestinal Absorption Using Sphere Exclusion and kNN QSAR Methods. *QSAR Comb. Sci.,* **2007**, *26*(5), 653 – 668.

[130] Iyer, M.; Tseng, Y.J.; Senese, C.L.; Liu, J.; Hopfinger, A.J., Prediction and mechanistic interpretation of human oral drug absorption using MI-QSAR analysis. *Mol Pharm,* **2007**, *4*(2), 218-231.

[131] Yan, A.; Wang, Z.; Cai, Z., Prediction of Human Intestinal Absorption by GA Feature Selection and Support Vector Machine Regression. *Int J Mol Sci,* **2008**, *9*(10), 1961-1976.

[132] Reynolds, D.P.; Lanevskij, K.; Japertas, P.; Didziapetris, R.; Petrauskas, A., Ionization-specific analysis of human intestinal absorption. *J Pharm Sci,* **2009**.

[133] Molnar, L.; Keseru, G.M., A Neural Network Based Virtual Screening of Cytochrome P450 3A4 Inhibitors. *Bioorg. Med. Chem. Lett.,* **2002**, *12*, 419-421.

[134] Ekins, S.; Berbaum, J.; Harrison, R.K., Generation and validation of rapid computational filters for cyp2d6 and cyp3a4. *Drug Metab Dispos,* **2003**, *31*(9), 1077-1080.

[135] Korolev, D.; Balakin, K.V.; Nikolsky, Y.; Kirillov, E.; Ivanenkov, Y.A.; Savchuk, N.P.; Ivashchenko, A.A.; Nikolskaya, T., Modeling of human cytochrome p450-mediated drug metabolism using unsupervised machine learning approach. *J Med Chem,* **2003**, *46*(17), 3631-3643.

[136] Merkwirth, C.; Mauser, H.; Schulz-Gasch, T.; Roche, O.; Stahl, M.; Lengauer, T., Ensemble Methods for Classification in Cheminformatics. *J Chem Inf Comput Sci,* **2004**, *44*(6), 1971-1978.

[137] Balakin, K.V.; Ekins, S.; Bugrim, A.; Ivanenkov, Y.A.; Korolev, D.; Nikolsky, Y.V.; Ivashchenko, A.A.; Savchuk, N.P.; Nikolskaya, T., Quantitative Structure-Metabolism Relationship Modeling of the Metabolic N-Dealkylation Reaction Rates. *Drug Metab Dispos,* **2004**, *32*(10), 1111-1120.

[138] Yap, C.W.; Chen, Y.Z., Quantitative Structure-Pharmacokinetic Relationships for Drug Distribution Properties by Using General Regression Neural Network. *J Pharm Sci,* **2005**, *94*(1), 153-168.

[139] Kriegl, J.M.; Eriksson, L.; Arnhold, T.; Beck, B.; Johansson, E.; Fox, T., Multivariate modeling of cytochrome P450 3A4 inhibition. *Eur J Pharm Sci,* **2005**, *24*(5), 451-463.

[140] Yap, C.W.; Xue, Y.; Li, Z.R.; Chen, Y.Z., Application of support vector machines to in silico prediction of cytochrome P450 enzyme substrates and inhibitors. *Curr Top Med Chem,* **2006**, *6*, 1593-1607.

[141] Jones, D.R.; Ekins, S.; Li, L.; Hall, S.D., Computational approaches that predict metabolic intermediate complex formation with CYP3A4 (+b5). *Drug Metab Dispos,* **2007**, *35*(9), 1466-1475.

[142] Terfloth, L.; Bienfait, B.; Gasteiger, J., Ligand-based models for the isoform specificity of cytochrome P450 3A4, 2D6, and 2C9 substrates. *J Chem Inf Model,* **2007**, *47*(4), 1688-1701.

[143] Bain, L.J.; McLachlan, J.B.; LeBlanc, G.A., Structure-activity relationships for xenobiotic transport substrates and inhibitory ligands of P-glycoprotein. *Environ Health Perspect,* **1997**, *105*(8), 812-818.

[144] Osterberg, T.; Norinder, U., Theoretical calculation and prediction of P-glycoprotein-interacting drugs using MolSurf parametrization and PLS statistics. *Eur J Pharm Sci,* **2000**, *10*(4), 295-303.

[145] Bakken, G.A.; Jurs, P.C., Classification of multidrug-resistance reversal agents using structure-based descriptors and linear discriminant analysis. *J Med Chem,* **2000**, *43*(23), 4534-4541.

[146] Dearden, J.C.; Cronin, M.T.D.; Al-Noobi, A.; Raevsky, O.A., QSAR modelling and prediction of P-glycoprotein-associated ATPase activity. **2001**.

[147] Osterberg, T.; Norinder, U., Prediction of drug transport processes using simple parameters and PLS statistics. The use of ACD/logP and ACD/ChemSketch descriptors. *Eur J Pharm Sci,* **2001**, *12*(3), 327-337.

[148] Ekins, S.; Kim, R.B.; Leake, B.F.; Dantzig, A.H.; Schuetz, E.G.; Lan, L.B.; Yasuda, K.; Shepard, R.L.; Winter, M.A.; Schuetz, J.D.; Wikel, J.H.; Wrighton, S.A., Three-dimensional quantitative structure-activity relationships of inhibitors of P-glycoprotein. *Mol Pharmacol,* **2002**, *61*(5), 964-973.

[149] Onishi, Y.; Hirano, H.; Nakata, K.; Oosumi, K.; Nagakura, M.; Tarui, S.; Ishikawa, K., Application of Three-Dimensional Quantitative Structure-Activity Relationships of P-Glycoprotein Inhibitors and Substrates. *Chem-Bio Informatics J,* **2003**, *3*(4), 175-193.

[150] Wang, R.B.; Kuo, C.L.; Lien, L.L.; Lien, E.J., Structure–activity relationship: analyses of p-glycoprotein substrates and inhibitors. *J. Clin. Pharm. Ther.,* **2003**, *28*, 203-228.

[151] Wang, Y.H.; Li, Y.; Yang, S.L.; Yang, L., Classification of substrates and inhibitors of P-glycoprotein using unsupervised machine learning approach. *J Chem Inf Model,* **2005**, *45*(3), 750-757.

[152] Cianchetta, G.; Singleton, R.W.; Zhang, M.; Wildgoose, M.; Giesing, D.; Fravolini, A.; Cruciani, G.; Vaz, R.J., A pharmacophore hypothesis for P-glycoprotein substrate recognition using GRIND-based 3D-QSAR. *J Med Chem,* **2005**, *48*(8), 2927-2935.

[153] Crivori, P.; Reinach, B.; Pezzetta, D.; Poggesi, I., Computational models for identifying potential P-glycoprotein substrates and inhibitors. *Mol Pharm,* **2006**, *3*(1), 33-44.

[154] Sheu, M.T.; Kao, Y.H.; Lin, Y.K.; Ho, H.O., A Quantitative Structure–Activity Relationship for the Modulation Effects of Flavonoids on P-Glycoprotein-Mediated Transport. . *Chem Pharm Bull,* **2010**, *58*(9), 1187-1194.

[155] Estrada, E.; Molina, E.; Nodarse, D.; Uriarte, E., Structural contributions of substrates to their binding to P-Glycoprotein. A TOPS-MODE approach. *Curr Pharm Des,* **2010**, *16*(24), 2676-2709.

[156] Yoshida, F.; Topliss, J.G., QSAR model for drug human oral bioavailability. *J Med Chem,* **2000**, *43*(13), 2575-2585.

[157] Turner, J.V.; Glass, B.D.; Agatonovic-Kustrin, S., Prediction of drug bioavailability based on molecular structure. *Anal Chim Acta,* **2003**, *485*, 89-102.

Received: February 20, 2011 Revised: August 17, 2011 Accepted: August 18, 2011