Expert Opinion on Drug Discovery

ISSN: 1746-0441 (Print) 1746-045X (Online) Journal homepage: http://www.tandfonline.com/loi/iedc20

Computational modeling of human oral bioavailability: what will be next?

Miguel Ángel Cabrera-Pérez & Hai Pham-The

To cite this article: Miguel Ángel Cabrera-Pérez & Hai Pham-The (2018): Computational modeling of human oral bioavailability: what will be next?, Expert Opinion on Drug Discovery, DOI: 10.1080/17460441.2018.1463988

To link to this article: https://doi.org/10.1080/17460441.2018.1463988

Published online: 17 Apr 2018.

EXPERT OPINION ON DRUG DISCOVERY, 2018
https://doi.org/10.1080/17460441.2018.1463988

Taylor & Francis
Taylor & Francis Group

REVIEW

Computational modeling of human oral bioavailability: what will be next?

Miguel Ángel Cabrera-Pérez [a,b,c] and Hai Pham-The [d]

[a]Unit of Modeling and Experimental Biopharmaceutics, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, Cuba; [b]Department of Pharmacy and Pharmaceutical Technology, University of Valencia, Burjassot, Spain; [c]Department of Engineering, Area of Pharmacy and Pharmaceutical Technology, Miguel Hernández University, Alicante, Spain; [d]Department of Pharmaceutical Chemistry, Hanoi University of Pharmacy, Hanoi, Vietnam

ABSTRACT

Introduction: The oral route is the most convenient way of administrating drugs. Therefore, accurate determination of oral bioavailability is paramount during drug discovery and development. Quantitative structure-property relationship (QSPR), rule-of-thumb (RoT) and physiologically based-pharmacokinetic (PBPK) approaches are promising alternatives to the early oral bioavailability prediction.

Areas covered: The authors give insight into the factors affecting bioavailability, the fundamental theoretical framework and the practical aspects of computational methods for predicting this property. They also give their perspectives on future computational models for estimating oral bioavailability.

Expert opinion: Oral bioavailability is a multi-factorial pharmacokinetic property with its accurate prediction challenging. For RoT and QSPR modeling, the reliability of datasets, the significance of molecular descriptor families and the diversity of chemometric tools used are important factors that define model predictability and interpretability. Likewise, for PBPK modeling the integrity of the pharmacokinetic data, the number of input parameters, the complexity of statistical analysis and the software packages used are relevant factors in bioavailability prediction. Although these approaches have been utilized independently, the tendency to use hybrid QSPR-PBPK approaches together with the exploration of ensemble and deep-learning systems for QSPR modeling of oral bioavailability has opened new avenues for development promising tools for oral bioavailability prediction.

ARTICLE HISTORY
Received 31 January 2018
Accepted 9 April 2018

KEYWORDS
Oral bioavailability; physiologically based-pharmacokinetic model (PBPK); quantitative structure-property relationship (QSPR); computational ADME; oral absorption; drug discovery; drug development

1. Introduction

Current pharmaceutical industry is facing a huge pressure of the high attrition rates in drug discovery and development [1]. Among the main bottlenecks responsible for this problem stand out the poor pharmacokinetic (PK) properties of drug candidates. It has been reported that 10% of oral drug failures in the development process are due to unfavorable absorption, distribution, metabolism, and excretion (ADME) properties [2].

Given the predominance and convenience of the patient for oral drug administration, the oral bioavailability of a new drug candidate remains one of the most important PK parameters. Having a low and highly variable intestinal absorption and bioavailability profiles could be the main factor that limits the success of a new drug candidate during the development process. Unfortunately, this situation is only manifested in the late stages of the development process, such as clinical phases. This failure would be much of money lost for the pharmaceutical industry every year, not to mention the great time and labor invested. Therefore, the early 'Go/No Go' decision based on drug candidate absorption profiles is prevalent to avoid such losses [3]. In order to support decision-making at this stage, computational or *in silico* methods have been established and widely applied to predict this absorption property in the early stages of drug discovery. On the basis of *in silico* predictions, the most promising candidates will be selected for the clinical trials [4]. Nevertheless, the computational prediction of human oral bioavailability is a great challenge because it involves numerous chemical and physiological processes such as chemical stability in the gastrointestinal (GI) tract, solubility and dissolution, formulation, intestinal permeability, first-pass effect in the gut, first-pass metabolism in the liver, among others [5].

Two main strategies are currently used for computationally modeling of human oral bioavailability: i) those based on chemical descriptors such as quantitative structure-property relationship (QSPR) and rule-of-thumb (RoT), and ii) those based on physiological features of human body components evolved in the absorption process (e.g. physiologically based pharmacokinetic (PBPK) modeling). In this study, we provided a systematic survey of the state-of-the-art and possible perspectives of these two approaches in predicting human oral bioavailability in drug discovery and development. We analyzed the main factors affecting bioavailability and the progression of computational modeling studies reported over the last two decades. Based on the fundamental theoretical framework of each method, its advantages, and limitations, we revealed some new lines for

CONTACT Miguel Ángel Cabrera-Pérez macabrera@uclv.edu.cu Unit of Modeling and Experimental Biopharmaceutics, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, Villa Clara, Cuba

Article Highlights

- Oral bioavailability is an important pharmacokinetic property for oral drugs and is one of the bottlenecks during the drug discovery and development stages.
- Two computational strategies have mainly been used to model oral bioavailability in humans, one based on the quantitative structure-property relationship (QSPR) and rule-of-thumb, and the other on physiologically based-pharmacokinetic (PBPK).
- Several factors are involved in the variables results obtained in the computational prediction of oral bioavailability with the QSPR approach: i) the database used, ii) type of molecular descriptor and iii) statistical method employed to obtain the models.
- Similar to QSPR approach, PBPK methods depend on the extent, quality and relevance of the input data used.
- The accurate computational prediction of oral bioavailability using QSPR and PBPK approaches is still an unsolved problem.

The use of hybrid QSPR-PBPK approaches along with the exploration of ensemble and deep-learning systems for QSPR modeling of oral bioavailability could be a new perspective in the development of promising tools for predicting oral bioavailability.

This box summarizes key points contained in the article.

developing computational tools with improved capability to predict oral drug bioavailability in human.

2. Understanding the oral bioavailability concept

According to the European Medicine Agency (EMA), oral drug bioavailability is defined as the *rate and extent to which an active moiety is absorbed from a pharmaceutical form and becomes available in the systemic circulation* [6]. In practice, the absolute bioavailability is measured by the ratio of the dose-corrected area under curve (AUC) of the oral route to that of the intravenous (IV) route.

In order to understand the complexity of the oral bioavailability parameter, the global biological process should be analyzed in stages (Figure 1).

The first step in obtaining high oral bioavailability is to achieve good oral absorption. The fraction of an orally administered dose that is absorbed in the gastrointestinal tract (F_a) is determined by fundamental physicochemical and biological properties such as dissolution rate, aqueous solubility and permeability (Figure 1). Gut bioavailability (F_g) is the fraction of orally administered drug that escapes to the first-pass gut wall extraction. Gut extraction of a drug (E_g) may be due to luminal degradation, efflux processes or intestinal metabolism. Hepatic bioavailability (F_h) is the fraction of drug that escapes to the first-pass hepatic metabolism (E_h) and biliary secretion. Thus, systemic bioavailability (F) is a product of the absorbed dose fraction and drug fractions that escapes from metabolism in the gastrointestinal tract, and the first-pass hepatic metabolism, respectively. The systemic bioavailability can be elucidated by the following equation: $F = Fa \times Fg \times Fh = Fa \times (1 - Eg) \times (1 - Eh)$

Although all these factors are known, their interaction is not always completely clear and only few of them have been used to predict oral bioavailability [7,8]. For example, Hou et al. [9] studied a dataset of 470 compounds and identified 36% of the data as highly metabolized molecules, and that oral bioavailability for 64% of the data was mainly controlled by absorption, suggesting that for a reliable prediction of oral bioavailability, the correct prediction of intestinal absorption and metabolism should be considered.

The measurement of the oral systemic bioavailability of a drug is influenced by other elements such as the concomitant administration of drug with or without food, or taken with other drugs, as well as the disease states. These factors may alter the drug absorption and the liver metabolism [10]. A summary of the most important factors affecting human oral bioavailability is shown in Table 1.

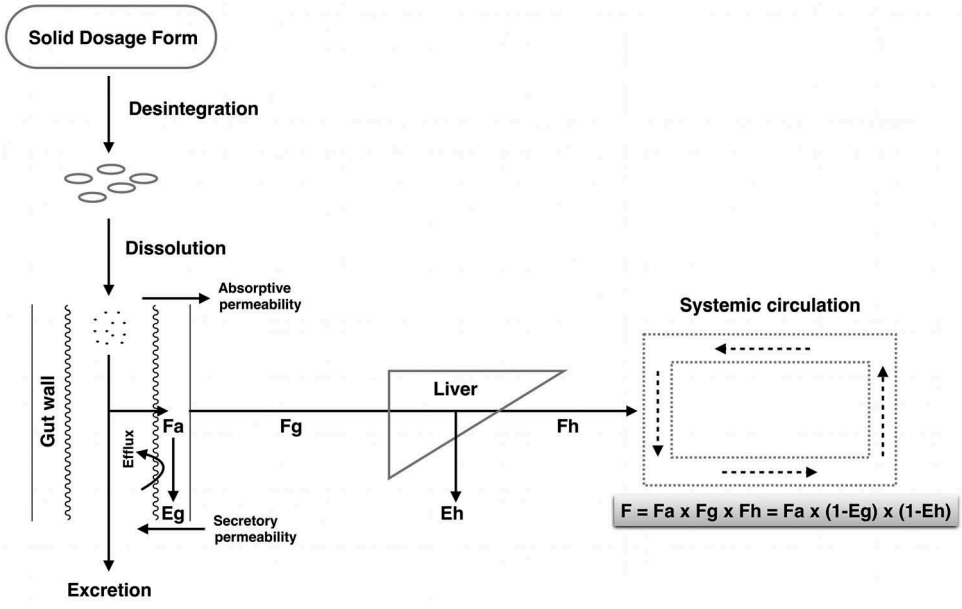

Figure 1. General representation of oral delivery process of drug to systemic circulation. Main processes involved in oral bioavailability are represented: F_a: fraction of an orally administered dose that is absorbed in the gastrointestinal tract; F_g: Gut bioavailability; E_g: Gut extraction of a drug; F_h: Hepatic bioavailability; E_h: Hepatic extraction of a drug, and F: Systemic bioavailability.

Table 1. Most important factors affecting human oral bioavailability.

Factors affecting oral bioavailability of drugs			
Physiological	Physicochemical	Dosage form and formulation	Others
Gastrointestinal pH interactions	Stability of drug in GI fluids	Disintegration rate	Drug
Stomach emptying rate/flow rate	Ionization constant	Dissolution rate	Food effects
Intestinal motility	Lipophilicity	Drug release mechanisms	Disease effects
Intestinal transit time	Solubility of drug in GI fluids	Excipient effects	
Passage of drugs across GI membrane	Crystallinity		
Unstirred water layer	Dissolution		
Bacterial microflora	Salt form		
Intestinal secretion	Particle size		
Intestinal efflux	Protein binding		
Luminal/hepatic metabolism	Polymorphism		
Age	Complexation		
Gender	Adsorption		
Transporter capacity	Charge		
	Diffusivity		

3. Non-physiological *in silico* models for prediction of human oral bioavailability

In order to develop computational models for predicting oral bioavailability in the early stages of drug discovery and development, researchers have followed different strategies such as the application of non-physiological models (RoT or QSPR models) to predict bioavailability-related biopharmaceutical properties, such as solubility, permeability, intestinal absorption and metabolism, and the use of PBPK to predict absorption and metabolism from *in vitro* or *in silico* inputs [11].

Non-physiological models of oral bioavailability based on QSPR approaches follow a general standard protocol: i) selection or curation of experimental dataset, ii) calculation of molecular descriptors, and iii) construction of the statistical model [12]. A more detailed description of these steps will be explained in the following sections.

3.1. Selection of the experimental bioavailability datasets

The predictability and robustness of the *in silico* bioavailability model is highly dependent of the experimental data quality [13]. For this reason, the quantity and quality of bioavailability data used to develop computational models play an important role in the chemical space coverage and performance of the model. In this regard, a detailed analysis about availability, size, and quality of data for published bioavailability datasets will be carried out. Table 2 describes the most relevant database/datasets used to develop computational models of human oral bioavailability.

As can be appreciated in Table 2, there are several datasets and databases published.

Unfortunately, in the public domain, the availability of experimental bioavailability data with quality and adequate structural diversity is limited. Almost all existing datasets are compilations of data derived from literature and only a few of them represent the results of controlled experimental assessment [26]. Usually, *in vivo* oral bioavailability values are collected from clinical trials of drugs or drug candidates, and these data may show significant variability from one source to another and are often biased toward compounds with high or moderate oral bioavailability, which influences the final predictive capacity of *in silico* models [27].

Public oral bioavailability datasets published before 2007 were generally much smaller than the larger nonpublic datasets, making it difficult to create highly reliable oral bioavailability models from these small datasets [5]. On the other hand, the lack of self-consistency of public databases, due to the fact that they mainly come from more than one source, produces a high level of variability in experimental data [10].

In 2007, Hou et al. [27] developed an extensive oral bioavailability dataset in human with 768 compounds from 185 published papers. This public dataset has been further extended and includes 1013 structurally diverse drugs and drug-like molecules, providing a good source for developing more reliable predictive models [28].

Another aspect to consider when using a published bioavailability dataset or when creating our own dataset, is the correct selection of the experimental data, since sometimes terms such as oral bioavailability and oral absorption are frequently exchanged in the literature [29,30].

However, access to good quality data of human bioavailability is probably the greatest barrier to improving the performance of current computational models.

In an effort to make high-quality PK data and predictive models available to the worldwide scientific community, freely available ADME databases have been published [31]. Among the main online databases with bioavailability data and other ADME endpoints are PKKB [32] and PK/DB [31].

3.2. Molecular descriptors

The molecular descriptor is the final result of transforming the chemical information of the molecule into a useful number, by means of mathematical or logical procedures [33]. Different classes of molecular descriptors have been employed to correlate chemical structure and ADME properties [34]. Although thousands of molecular descriptors are available, their selection must be based on basic knowledge of the property to be modeled, in order to develop successful prediction models [10]. In the prediction of oral bioavailability, descriptors with 1D, 2D, and 3D structural information have proven to be useful in deriving robust and predictive models [26].

One of the most important properties of the final computational model is its mechanistic interpretation, which begins with the number and the nature of the molecular descriptors used in the model [35]. For example, 1D-descriptors permit an easy mechanism interpretation of bioavailability models, providing information about atoms, bond composition, molecular weight, etc. However, almost all models using this type of descriptor have limited accuracy, robustness, and predictive capacity. Nevertheless, this kind of descriptor has been used to estimate oral bioavailability

Table 2. Datasets or database with oral bioavailability information.

Reference	Authors (year)	Availability	Size (No. of chemicals)	Comments
[29]	Sietsema et al. (1989)	Public	409 (14 prodrugs)	No information regarding formulations and/or dose, not all the compounds listed have both animal and human F data
[14]	Hirono et al. (1994)	Public	188	Data collected from literature
[15]	Bertz and Granneman (1997)	Public	315	Although is a metabolism database, values of oral bioavailability are reported
[47]	Yoshida and Topliss (2000)	Public	232	Bioavailability value of each drug was represented as a rating according to the ranges (four classes)
[48]	Andrews et al. (2000)	In-house	591	Data obtained from literature and an internal database (compounds collected from GlaxoWellcome)
[36]	Veber et al. (2002)	In-house	1100	Rat bioavailability data. Included a data of human bioavailability (277 compounds) taken from Goodman & Gilmans (editions 8 and 10). Data collected from GSK.
[16]	Pintore et al. (2003)	Public	507	Included bioavailability data from references [15,29] and Goodman & Gilmans (9th edition). Data divided in four classes.
[17]	Turner et al. (2003)	Public	169	Data for all compounds were taken from literature.
[18]	Turner et al. (2004)	Public	167	Data for all compounds were taken from literature.
[19]	Wang et al. (2006)	In-house	577	Included bioavailability data mainly from references [29] and Goodman & Gilmans (9 and 10th editions). Other data collected from Encysive
[27]	Hou et al. (2007)	Public	768	UCSD ADME database. Bioavailability data, collected from 185 papers, is included in the database http://modem.ucsd.edu/adme
[20]	Moda et al. (2008)	Public	302	Bioavailability data collected from references [29], Goodman & Gilmans (10th edition) and Physicians' Desk Reference (PDR)
[49]	Wang et al. (2008)	Public	772	Same dataset published by Hou et al. (2007)
[21]	Ma et al. (2008)	Public	766	Same dataset published by Hou et al. (2007). Bioavailability data divided in two classes.
[28]	Tian et al. (2011)	Public	1013	Update previous Hou et al. bioavailability dataset. http://modem.ucsd.edu/adme
[74]	Ahmed et al. (2012)	Public	969	Dataset collected from Hou et al. (2007) and Moda et al. (2008)
[22]	Xu et al. (2012)	Public	805	Dataset collected from Hou et al. (2007)
[23]	Olivares-Morales et al. (2014)	Public	184	Collected from reference [24]. For this set of compounds is reported bioavailability in human and preclinical species.
[75]	Kim et al. (2014)	Public	995	Bioavailability data was classified as low (F% < 50) and high (F% ≥ 50).
[31]	PK/DB (2008)	Public	660	A global ADME database that include oral bioavailability data. http://www.pkdb.ifsc.usp.br
[32]	PKKB (2012)	Public	992	PharmacoKinetics Knowledge Base database (PKKB). Include a bioavailability dataset. http://cadd.suda.edu.cn/admet
[25]	Wombat-PK	Commercial	1125	Database containing > 6500 clinical pharmacokinetic measurements. http://www.sunsetmolecular.com/

by property-based rules [28]. Veber et al. [36] proposed a simple rule based on the number of rotatable bonds and the polar surface area or hydrogen bond count to predict oral bioavailability in rats. Martin [37] also proposed a scoring scheme based on molecular properties, the polar surface area, the Lipinski rule of five, and the molecular charged state to predict rat bioavailability. Although several efforts have been made to predict oral bioavailability, Hou et al. [28] found that no effective rule is able to predict oral bioavailability because these rules cannot explain the relevance of metabolic processes on this PK property. In this sense, Varma et al. [38] defined a physicochemical space for optimum human oral bioavailability, considering the combination of the effects of properties, such as molecular weight, ionization state, lipophilicity, polar descriptors, and free rotatable bonds on F_a and first-pass elimination (F_g and F_h). This study introduced a new perspective of using molecular properties, expressed in 1D-descriptors, to predict oral bioavailability.

The fast calculation speed for large datasets, as well as their sometimes-feasible mechanistic interpretation, makes 2D-molecular descriptors (e.g. constitutional, fragment, functional group-based, topological) a good alternative for use in predicting oral bioavailability. Among them, the electrotopological state indices (E-state) [39] and the TOPS-MODE [40] descriptors are good options with respect to interpretability, in the first case in terms of hydrogen bonding and in the second for their mechanistic interpretation at the bonding level and their capacity to generate structural alerts [26].

The 3D molecular descriptors (e.g. WHIM, Volsurf) provide more structural information compared with lower-dimensional descriptors, but the main problem is the selection of the correct 3D conformation. Several types of 3D descriptors have been used to predict bioavailability properties, but their selection is influenced by the complexity and calculation speed [26].

Some experimental PK properties such as bioavailability in animals, volume of distribution, clearance, area under the plasma concentration-time curve, etc., could be potential descriptors to predict oral bioavailability in humans. However, the use of descriptors based on experimental properties limits the number of compounds to be used in the model, and decreases its screening capacity in the initial stages of drug discovery and development where many of these properties have not been determined.

Different softwares have been developed to compute a large variety of molecular descriptors [41]. A summary of the main software employed can be appreciated in Table 3.

Table 3. A summary of software programs and web tools developed to calculate molecular descriptors and to predict ADME properties related with the oral absorption and/or bioavailability (QSPR software).

Software	Website	Description
Molecular descriptor		
DRAGON	https://chm.kode-solutions.net/products_dragon.php	Software for calculation of all types of molecular descriptors (5270)
E-Dragon	http://www.vcclab.org/lab/edragon/	Online version of DRAGON, an application for the calculation of molecular descriptors
CODESSA	http://www.codessa-pro.com	Calculation of over 500 types of molecular descriptors
Molconn-Z	http://www.edusoft-lc.com/molconn/	Software for generation molecular descriptors for QSAR analysis
Molecular Operating Environment (MOE)	http://www.chemcomp.com/software-chem.htm	This software calculates over 600 molecular descriptors and includes tools for the creation of QSAR/QSPR models
Mold2	http://www.fda.gov/ScienceResearch/BioinformaticsTools/Mold2	This software calculates over 700 molecular descriptors
Chemistry Development Kid (CDK)	http://cdk.sourceforge.net	A Java library for chemoinformatic applications. It includes the generation of 260 types of molecular descriptors
MOLGEN	http://molgen.de	Web service for calculating 708 molecular descriptors
PreADMET	http://preadmet.bmdrc.org	Calculates more than 2000 descriptors, with prediction of ADME/T and drug-likeness properties
ADMEWORKS Model Builder	http://www.fujitsu.com/jp/group/kyushu/en/solutions/industry/lifescience/admeworks/	A tool for building mathematical models to be used in the prediction of chemical and biological properties
VolSurf+	http://www.moldiscovery.com/software/vsplus/	Calculate ADME properties and create predictive ADME models
PaDEL Descriptors	http://www.yapcwsoft.com/dd/padeldescriptor/	Free software to calculate molecular descriptors and fingerprints
MOLE db	http://michem.disat.unimib.it/mole_db/	Free online database comprise of 1124 molecular descriptors
QSPR for ADME		
Impac-F	http://www.pharmainformatic.com/html/impact-f.html	Expert system to predict human oral bioavailability of drug candidates. Several QSAR models compose this system
ADMET Predictor	http://www.simulations-plus.com/software/admetpredictor/	This software estimates a number of ADMET properties from molecular structure
ACD/ADME Suite	http://acd-adme-suite.software.informer.com/	Predicts ADME properties (e.g. bioavailability) from chemical structure
Discovery Studio/ ADMET software	http://accelrys.com/products/collaborative-science/biovia-discovery-studio/qsar-admet-and-predictive-toxicology.html	ADMET properties can be calculated from collections of molecules
SwissADME	http://www.swissadme.ch/	Web tool to predict ADMET properties
admetSAR	http://lmmd.ecust.edu.cn/admetsar1/home/	A compressive source and free tool for assessment of chemical ADMET properties
ACD/percepta predictors	http://www.acdlabs.com/products/percepta/predictors.php	Prediction of ADME/T and physicochemical properties
Chembench	http://chembench.mml.unc.edu	Chemoinformatics support by integrating robust model builders, generators of descriptors, property predictors, etc.
KNIME	http://www.knime.org	Graphical workbench for the entire analysis process, including plug-ins for descriptor generation and creation of QSAR models
ADMEWORKS Predictor	http://www.fqs.pl/en/chemistry/products/admeworks-predictor	A virtual screening system intended for simultaneous evaluation of the ADMET properties of compounds
StarDrop	http://www.optibrium.com	QSAR modeling, data analysis, and ADME/T prediction
FAF-Drugs 4	http://fafdrugs4.mti.univ-paris-diderot.fr	Free software for *in silico* ADMET filtering

3.3. Prediction models

Since the late 1990s and early 2000s, developing computational models to predict early intestinal absorption and oral bioavailability has received significant attention and become the hot issues during the drug research and development (R&D) processes, with the main objective of increasing the success rate of new chemical entities (NCEs) [10,42].

Up to date, two modeling approaches have been widely explored and have made great advancement, such as the QSPR and RoT methods. The first focuses mainly on the development of mathematical models that express the correlation between molecular descriptors and the extent of oral bioavailability of compounds. Meanwhile, the latter refers to the simplest models, which are based on overall trends that relate physicochemical properties to oral bioavailability. For the development of simple rules, an oral bioavailability cutoff value that makes sense after the descriptor calculation must be identified. One of the most common thresholds is the 20% bioavailability value, which currently represents an acceptable criterion for screening potential lead during drug development processes [27,36].

In comparison with QSPR, RoT approach only gives a round prediction of absorption level of compounds and exhibits lower precision when applied for certain chemical families, such as peptides and peptidomimetics, natural products, macrocyclic compounds, etc. [43,44]. However, given the simplicity and transparent interpretation characteristics, RoT continues to be a more popular tool compared to QSPR models which use complex chemoinformatic techniques to obtain high performance models [45]. The following part is a brief summary of the progression of QSPR and RoT models development and application for oral bioavailability prediction in the last two decades.

According to the rule-based approaches, several rules have been constructed by proposing cutoff values of physicochemical properties that allow defining specific bioavailability classes. In a paper released in 2002, Veber et al. [36] analyzed over 1100 drug candidates studied at GlaxoSmithKline and identified some common features of compounds with rat oral bioavailability of 20% or greater, which are as follows: i) number of rotatable bonds (RBN) ≤ 10, and ii) polar surface area (TPSA(tot)) ≤ 140 $Å^2$ or the sum of H-bond acceptors and H-bond donors ≤ 12. Later, Lu et al. [46] examined the

accuracy of the Veber's rule for 434 Pharmacia compounds and found a lower accuracy of the Veber's rule as compared to the original report. Concerning the false positive rate in early screening, Hou et al. [27] applied the Veber's rule for a database of human oral bioavailability for 768 chemical compounds. The analysis showed that the rule was unable to identify compounds with poor oral bioavailability. Other interesting rules are bioavailability scores proposed by Martin [37] on the basis of 553 compounds studied at Abbott Laboratories. The scores are based on the calculation of PSA, rule-of-five compliance, and the predominant charge of compounds at biological pH. Some recent studies have focused on mapping physicochemical space to optimize oral bioavailability of drugs. Varma et al. [38] separated the three main components of oral bioavailability: absorption (F_a), first-pass gut metabolism (F_g), and first-pass hepatic extraction (F_h), and their analysis evidenced that intestinal absorption and first-pass elimination are associated with a different physicochemical space. Although the observations made in this study are general trends, in our opinion the interrelationship of these physicochemical properties and bioavailability should be considered for compounds that are substrates of CYP3A4 and Pgp.

In contrast to the rules, since 2000s numerous QSPR models have been developed to predict the oral bioavailability of drugs in humans and rodents, beginning with the initial efforts of Yoshida and Topliss [47], and those of Andrews et al. [48] (see Table 4). In general, most of computational models attempted to quantitatively estimate the values of F. Diverse statistical and intelligence artificial techniques have been applied; however the performance of these models appears to be rather modest and data-dependent. Clearly, the low accuracy and high variability of current bioavailability measurements make the development of predictive QSPR models a thorny task. In addition, for different types of compounds, PK factors such as absorption, distribution, and metabolism contribute with different weight to the final bioavailability extent. In this context, some authors have attempted to improve the prediction of bioavailability using human oral administration data (e.g. human intestinal absorption) and animal PK data (e.g. the distribution volume of the terminal phase and the elimination rate constant) as descriptors. Examples of this are the reports by Wang et al. [49] and Imawaka et al. [50].

However, in the early stages of drug design, it is difficult to measure these experimental descriptors, as only a small number of compounds are available. To overcome this situation, we have recently proposed the consensus use of numerous QSPR models that describe all the specific events that a compound faces until it reaches systemic circulation to produce a therapeutic effect [4]. These events include, but are not limited to, the ability of a molecule to dissolve in water, permeate through the intestinal wall, and escape from the intestinal and hepatic metabolism. Once the results of these models are obtained, they can be used as new theoretical descriptors useful for the development of the final bioavailability QSPR model [4]. In conclusion, as Wang and Hou [51] have pointed out, there is still a long way to go to successfully develop chemoinformatic tools that can reliably predict human oral bioavailability.

3.4. Softwares

In almost all published QSPR models for predicting oral bioavailability, the authors use various statistical modeling methods to correlate molecular descriptors with the ADME property (see Table 4). Some of these models have subsequently been implemented in specific software. On the other hand, a large number of software programs are available to predict oral absorption or other PK properties. Table 3 also summarizes some commercial programs applied in the prediction of absorption properties. Many of these ADME tools are applied during the early stage of the drug development process to identify compounds with poor bioavailability values, with significant savings in R&D costs.

Table 4. Summary of up-to-date QSPR models developed for oral bioavailability prediction.

Reference	Model type	Modeling method	Model size	Model performance*
[47]	Classification	ORMUCS[a]	18 physicochemical descriptors and fragments	N_{train} = 232, N_{test} = 40, R^2 = 0.71, Q^2_{LOO} = 0.67, Q^2_{test} = 0.60
[48]	Regression	MLR[b]	85 fragments	N = 591, R^2 = 0.71, Q^2_{LOO} = 0.63, Q^2_{LGO} = 0.58 (80/20 splits)
[17]	Regression	MLR	8 descriptors	N_{train} = 159, N_{test} = 10, R^2 = 0.35, Q^2_{LOO} = 0.25, Q^2_{test} = 0.72
[16]	Classification	AFP[c]	10 descriptors	N_1 = 272, R^2 = 0.82, Q^2_{test} = 0.40, $Q^2_{validation}$ = 0.75 N_2 = 432, R^2 = 0.70, Q^2_{test} = 0.64, $Q^2_{validation}$ = 0.68
[18]	Regression	ANN[d]	10 descriptors	N_{train} = 137, N_{test} = 15, $N_{validation}$ = 15, R^2 = 0.74, Q^2_{test} = 0.68, $Q^2_{validation}$ = 0.90
[19]	Regression	GA[e]-MLR	8 descriptors, 42 fragments	N = 577, R^2 = 0.55, Q^2_{LGO} = 0.42 (90/10 splits)
[20]	Regression	Hologram-QSAR	8 components	N_{train} = 250, N_{test} = 52, R^2 = 0.93, Q^2_{test} = 0.70
[21]	Classification	GA–CG–SVM[f]	25 Cerius[2] descriptors	N_{train} = 690, N_{test} = 76, $Q^2_{5\text{-fold CV}}$ = 0.80, Q^2_{test} = 0.86
[28]	Regression	GA- MLR	60–100 descriptors, fingerprints	N_{train} = 996, N_{test} = 80, best models (6, 10 and 11): R^2 = 0.73–0.79, Q^2_{LOO} = 0.68–0.72, Q^2_{test} = 0.60–68
[22]	Regression	Stepwise-MLR, PLS[g], SVM	7–21 descriptors, 1536 descriptors (SVM^T)	Set 1: N_{train} = 156, N_{test} = 36, best model (SVR^T): R^2 = 0.84, Q^2_{test} = 0.73 Set 2: N_{train} = 122, N_{test} = 27, best model (SVR): R^2 = 0.75, Q^2_{test} = 0.63 Set 3: N_{train} = 180, N_{test} = 44, best model (SVR): R^2 = 0.78, Q^2_{test} = 0.80 Set 4: N_{train} = 197, N_{test} = 43, best model (PLS): R^2 = 0.83, Q^2_{test} = 0.60
[74]	Classification	Logistic classifier	47 descriptors	N = 969, $Q^2_{overall}$ = 0.71, Precision = 0.73–0.75
[75]	Classification/ Regression	RF[h], SVM, kNN, CASE Ultra	1597 (Dragon) and 186 (MOE) descriptors	N_1 = 995, $R^2_{consensus}$ = 0.28, MAE = 24; N_2 = 362, $R^2_{consensus + MDRP/MDRI}$ = 0.40, MAE = 21 $Q^2_{validation}$ = 0.76

[a]Ordered multicategorical classification method using the simplex technique (ORMUCS), [b]Multiple Linear Regression, [c]Adaptive fuzzy partition, [d]Artificial neural network, [e]Genetic Algorithm, [f]Support vector machine, [g]Partial Least Square regression, [h]Random Fores

4. Physiological *in silico* models for prediction of human oral bioavailability

Although several QSPR and RoT models have been developed for bioavailability, their predictive power is highly variable and this is mainly due to the fact that this type of model has a limited ability to explain more complex and dynamic phenomena, such as pH-dependent oral absorption and food effects.

PBPK models are mathematical models that integrate knowledge of physiological processes (e.g. GI transit time, organ blood flows) with the physicochemical properties of compounds (e.g. solubility, lipophilicity) to predict or simulate complex biological properties [52]. In this sense, the input parameters of the *in silico, in vitro*, and *in vivo* experiments can be combined to predict the plasma and tissue concentration time profile [53].

An example of a generic whole-body PBPK model is described in Figure 2, where the figures are compartments representing tissues and organs from which blood flows carry a drug in and out [54].

4.1. Methodology

Figure 3 shows a modeling and simulation strategy to show how to use currently available PBPK models to predict PK properties during the drug discovery and development process.

In order to develop a PBPK model for predicting or simulating PK profiles (concentration-time), four general steps should be considered: i) the representation of the model, where the relevant organs or tissues are selected for inclusion in the model, incorporating the exposure, and metabolic pathways of the drug; ii) the parameterization of the model, where physiological, physicochemical, and biological parameters are included in the equations of the model; iii) model simulation, which involves predicting drug absorption and disposition in a defined administration scenario by numerically integrating the set of differential mass balance equations; and iv) model validation, where predicted results are compared with external experimental data, including parameter sensitivity analysis and uncertainty analysis [55].

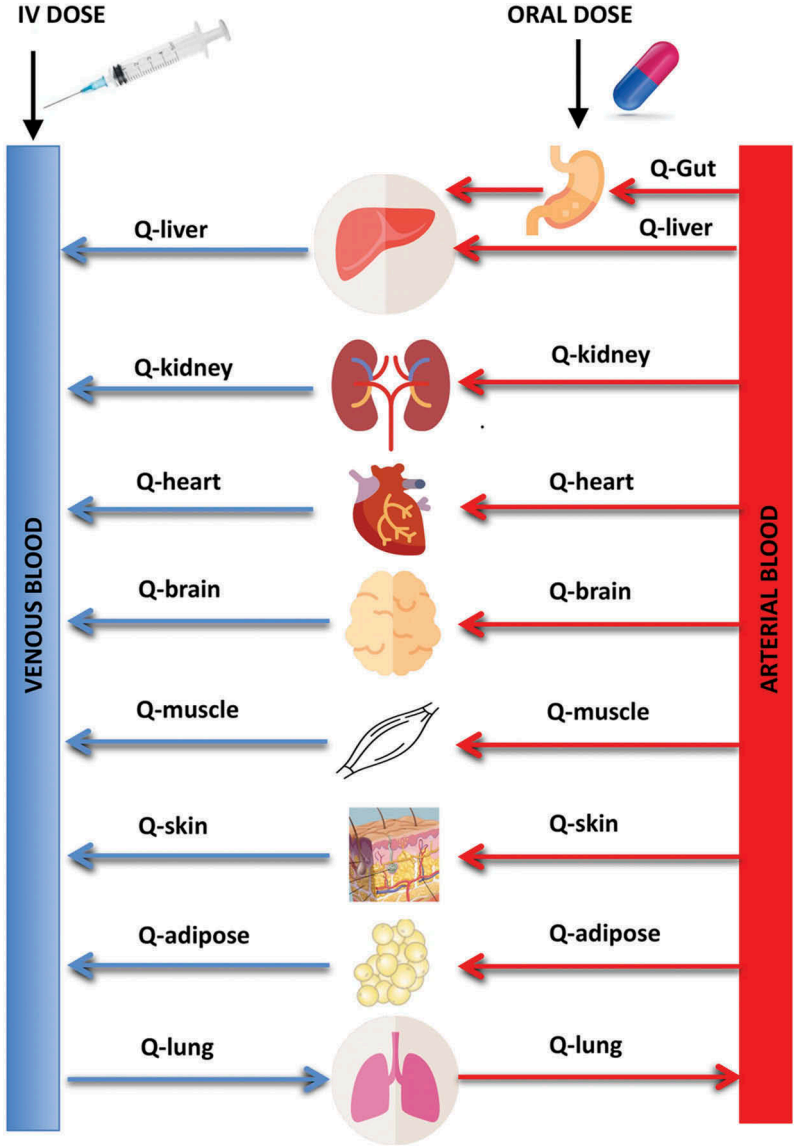

Figure 2. General representation of a physiologically-based pharmacokinetic model (PBPK) where compartments (figures) represent tissues and organs connected by blood flows (Q). An intravenous (IV) administration and an oral (O) administration are illustrated.

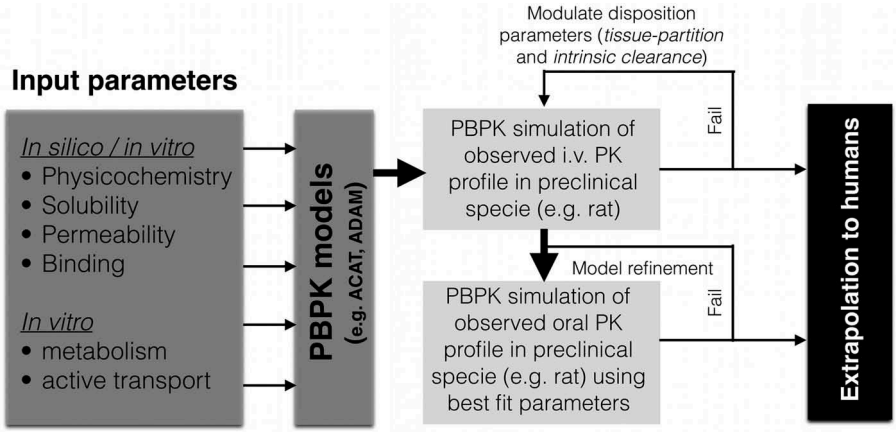

Figure 3. Schematic illustration of PBPK modeling and simulation strategy using a combination of modeling and experimental data.

Considering the PBPK approach as a predictive tool, the selection of reliable input parameters is very important, mainly for experimental measurements of tissue-partition coefficients and clearance, where *in vitro* measurements of intrinsic clearance scarcely match up to the *in vivo* value [56]. Initially, the input absorption parameters can be determined by QSPR methods, and then the solubility and permeability measurements can be replaced by experimental values (e.g. Caco-2, Ussing chamber, PAMPA). In order to overcome distribution and metabolism limitations, IV PK profiles can be used to extract metabolism and distribution parameters for better interpretation of more complex oral PK profiles [57]. These parameters are determined, keeping the rest of input parameters constant, until the best fit to the observed IV profiles is obtained. To make this comparison with the *in vivo* concentration-time profile, the rat is the model of choice because it is the most commonly used animal species for the PK characterization of new drug candidates. Furthermore, almost all generic PBPK models include suitable distribution and metabolic data for this species.

If the simulate plasma curve is consistent with experimental data in rats, different information about the underlying PK processes can be obtained and IV PKs can be simulated in humans. In case of inconsistency between the predicted PK profile and the observed PK profile, it should be necessary to refine the model before predicting human PK [56]. Once the best-fit distribution and metabolism parameters have been obtained, an oral simulation of the PK profile can be carried out.

This iterative feedback procedure is essential to facilitate a better mechanistic understanding of the absorption properties of new drugs or potential clinical candidates.

The combination of *in silico* and experimental methods to predict and simulate the absorption process has the potential to improve the selection and optimization of new drug candidates during the drug discovery and development stages.

4.2. Application of PBPK models for the prediction of oral absorption and bioavailability

The PBPK models have been widely applied during the drug discovery and development stages [58]. In recent years, the use of PBPK models has increased in the prediction of oral drug absorption, bioavailability, and the time course of the fraction of dose reaching systemic circulation [59]. According to Huang et al. [60], about 9% of the IND/NDA submissions applied to PBPK modeling and simulation are related to absorption and bioavailability properties.

Since the rate and extent of oral bioavailability is a complex process, consisting of a combination of pharmaceutical, physiological, and metabolic transport factors, the prospective prediction of this property remains a challenge, both during the drug discovery process and in the development of drug candidates.

To this end, physiologically based models have developed several mechanistic absorption models to describe the intestinal transit, such as: i)tank models, in which the GI tract is considered a single and well-stirred compartment; ii) compartmental transit models, in which the intestinal tract is modeled by a series of compartments; and iii) dispersion models, in which a continuous tube including transport and dispersion is used [61]. These models have been included in some commercial software packages and in-house developments. A detailed description of these software programs will be discussed in the next section.

The first attempts to predict oral absorption using PBPK models were described by Yu and Amidon [62,63], who developed a compartmental model of absorption and transit (CAT) to predict the fraction of absorbed dose (F_a) of different drugs. These integrated models include seven transit compartments, representing different anatomical regions of the small intestine. This version of the CAT model does not consider several properties that affect drug absorption, such as rate of dissolution, pH dependence on drug solubility, absorption in the stomach and/or colon, first-pass metabolism and drug degradation in the intestine and liver, and other factors within the intestinal tract, which limit its application in predicting absorption for compounds with low solubility or permeability. In order to solve this situation, an advanced CAT model (ACAT) was proposed, which describes the human GI tract with nine segments, from the stomach to the colon, where all the processes of the intestinal tract are considered. This model has been successfully applied to predict different aspects of drug absorption and bioavailability [11]. In 2005, Parrot et al. [64] predicted oral bioavailability in rats of 64 compounds, using

the ACAT model (e.g. GastroPlus™) and combining *in silico, in vitro*, and *in vivo* data.

Insufficiently predicted bioavailability results were obtained because the model does not clearly capture solubility problems. More recently, Paixao et al. [65] used a variation of the ACAT model and combined *in vitro* and *in silico* data to predict oral bioavailability. Good results were achieved when *in vitro* permeability (P_{app}) and intrinsic clearance (Cl_{int}) data were used, while predictions were lower when only *in silico* data were utilized.

In order to make a more accurate prediction of the oral absorption of the drug, an update version of the ACAT model was carried out, including the secretion and absorption of fluids, and incorporating an elaborate representation of the intestinal mucosa (PK-Sim®) [66]. The suitability of the model was demonstrated with a data set of 111 passively absorbed drugs with limited permeability. An extended version of the previous PBPK absorption model was developed, including experimental *in vitro* dissolution profiles, expressed as Weibull function fits [67]. In this case, the PBPK model is used as complementary information to the *in vitro* dissolution test to accurately predict the oral absorption profiles of different dosage forms.

The advanced, dissolution, absorption, and metabolism (ADAM) model has a similar structure to the CAT and ACAT models but offers the possibility of simulating populations and their variability [68].

Darwich et al. [69] used the ADAM model to analyze the interaction between the parameters of the model that affect dissolution, absorption, transport, and metabolism in bioavailability. The results of this work evidenced the expected high relationship between the metabolizing enzymes and the efflux transporters, using a realistic range of parameter values taken from literature.

Cai et al. [70] used the PK express model, a physiological model for predicting metabolism, and the iDEA absorption module for predicting human oral bioavailability. For about 80% of the data, there was a correspondence between the predicted bioavailability and the known clinical values of this property.

Considering that almost global PBPK models for predicting bioavailability from *in silico* and experimental inputs have not achieved the <2-fold average error to guide the lead optimization process, Daga et al. [71] developed local PBPK models (GastroPlus) to predict the bioavailability of three individual chemical series. The authors developed a local QSPR for a numerically fitted effective intrinsic clearance, and the rest of software inputs were calculated from the structure alone [71]. Successful predictions were obtained only from the chemical structure, suggesting the use of these models to predict the bioavailability of analogs prior to synthesis.

PBPK models have also been used to explain the differences in bioavailability between the problems of immediate-release (IR) and modified-release (MR) formulations for a specific drug. In this sense, Olivares-Morales et al. [72] combined *in vivo, in vitro*, and *in silico* data to predict the oral bioavailability of Oxybutynin in the OROS® formulation compared to the IR formulation.

4.3. Software

There are several commercial software packages on the market for physiology-based modeling of PK processes (see Table 5). Many of them can be used at all stages of drug discovery and development, mainly in preclinical and clinical development [73].

Two general categories of software are currently used for PBPK modeling and simulation. The first is 'open' software, which is not designed for PBPK modeling and full functionality requires a user experience in the programming language and in the modeling process itself.

The second case is the 'designed' software, which has been developed specifically for PBPK modeling and whose 'user-friendly' feature explains its widespread use [53].

5. Conclusion

In this review, a comprehensive analysis of the progress of computational models in predicting oral bioavailability was conducted. This PK property is very complex and depends on several physiological and physicochemical factors. Among the widely used *in silico* methods for predicting oral bioavailability have been the QSPR and rule-based approaches. However, the predictability of the main published bioavailability models is limited principally due to the lack of reliable experimental databases, the inability of molecular descriptors to explain metabolic process, as well as some modeling approaches used. In this sense, a new strategy is proposed, such as the consensus model approach to develop better QSPR models of bioavailability.

On the other hand, PBPK models for predicting human bioavailability are a powerful tool for pharmaceutical researchers and scientists in drug discovery and development. As with the QSPR approach, the final results achieved with PBPK methods depend on the extent, quality, and relevance of the input data used. In this sense, the input data must be generated in such a way that it better reflects the situation *in vivo*.

To date, both modeling methods have been used independently to predict bioavailability, however, the tendency to use hybrid QSPR-PBPK approaches together with the exploration of ensemble and deep-learning systems for QSPR modeling has opened up new perspectives in the development of promising tools for predicting oral bioavailability.

6. Expert opinion

Oral bioavailability is a very complex PK property and its prediction has always been a challenge for academics and industry researchers. Among the *in silico* models for predicting oral drug bioavailability, QSPR and RoT approaches have been widely used. There are several factors involved in the variable computational prediction of oral bioavailability with these methods, and most of them are associated with the database used, as well as the type of molecular descriptor or statistical method utilized to obtain the models.

Since the introduction of QSPR models to predict bioavailability, the lack of a comprehensive and reliable experimental database has made it difficult to develop good predictive

Table 5. Commonly used software and tools currently used to perform PBPK modeling and simulation.

Developer	Software	Category	Website	Description
The Mathworks Inc.	SimBiology/ MATLAB®	Open	https://www.mathworks.com/products/simbiology.html	A programmatic tool to model, simulate, and analyze dynamic systems, focusing on pharmacokinetic/pharmacodynamic (PK/PD) and systems biology applications.
University of Southern California	ADAPT 5	Open	https://bmsr.usc.edu/software/adapt/	A computational modeling platform developed for pharmacokinetic and pharmacodynamic applications.
University of California	Berkeley-Madonna	Open	http://www.berkeleymadonna.com	The fastest, most convenient, and general purpose differential equation solver available today.
The Epsilon Group (TEG)	SAAM II	Open	https://tegvirginia.com/software/saam-ii/	A modeling, simulation, and analysis software package which supports the development and statistical calibration of compartmental models in biological, metabolic, and pharmaceutical systems.
Cyprotex Ltd	Cloe PK™ Cloe Predict	Designed	http://www.cyprotex.com/cloepredict/	Pharmacokinetic prediction using PBPK. Software designed for prediction of rat and human PK at the earliest stages of drug discovery based on early preclinical data.
Simulations Plus Inc.	GastroPlus	Designed	http://www.simulations-plus.com	Software based on ACAT model, which includes some modifications of the CAT models. This model brings good performance of formulation and drug-related processes such as release, dissolution, precipitation, degradation, and absorption. It has an optimization module for drug metabolism, influx and efflux transport in the enterocytes.
Computing in Technology	MEDICI-PK™	Designed	http://www.cit-wulkow.de/	Software for PBPK modeling of drug and multiple metabolites and interactions in humans and preclinical species.
Bayer Technologies Services	PK-Sim™	Designed	http://www.systems-biology.com	In this software, the gastrointestinal tract is modeled as a continuous tube with spatially varying properties. It is used for oral absorption and PBPK modeling in preclinical species and humans that can simulate physiological variability in response.
Simcyp Ltd	Simcyp® Simulator	Designed	http://www.simcyp.com	It is based on ADAM model, which is divided into nine compartments from the stomach to the colon. The absorption process in each compartment is described as a function of release, dissolution, precipitation, luminal degradation, permeability, metabolism, transport, and transit to one segment to another. This permits a more complete prediction of oral bioavailability and to evaluate the impact of inter- and intra-subject variability due to physiological, pathological, and genetic factors. It is a clinical trial simulator for PK and DDI studies with PBPK modeling capabilities and algorithms for scaling *in vitro* data.
Fujitsu Kyushu Systems	ADMEWORKS DDI Simulator		http://www.fqs.pl/en/chemistry/products/admeworks-ddi-simulator	Predict the extent of drug-drug interactions arising from co-administration of drugs, an important study in drug development, through time course simulation of the concentrations of each drug in the body using PBPK mathematical models.
Intellipharm	Intellipharm®PK		www.intellipharm.com	This software simulates drug dissolution, absorption and pharmacokinetics (PK). It is based on mixing tank model. The software permits to define time-varying solubility, absorption constant and volume, bring a good choice to study the true movement along the GI tract. The original absorption model is coupled to classical compartment PK models, simulating plasma concentration *vs* time profiles and assessing the effect of absorption rate on drug exposure.
Isee System	Stella™		https://www.iseesystems.com/	A modeling software package that diagrams, charts, and uses animation help visual learners discover relationships between variables and helps simplify model building

models. This is true for *in vivo* values of human intestinal absorption and bioavailability, which are taken from reported data on marketed drugs or drugs in clinical trials [68]. Although some QSPR models have been developed with reliable and extensive databases, many of them are not available to the scientific community, which limits the possibility of obtaining new and better bioavailability predictive models [36,48]. Recently, some extensive bioavailability databases have been published, which will increase the quality and applicability of predictive models in the early stages of drug discovery and development [28,74,75].

Not only the size of the dataset is important, but also the type of data used. Many of the bioavailability data is based on single concentration determinations, which limits the influence of the transport mechanism and metabolism on the final bioavailability value [65].

Another factor affecting the predictability of *in silico* bioavailability models is related to the type of molecular descriptor used, as most of them do not considered the contribution of intestinal and hepatic metabolism to this property [51]. Some authors have suggested the introduction of new molecular rules or descriptors to improve the prediction of *in silico* bioavailability models [9,27]. Although some attempts have been made [38], in our opinion, a better characterization of factors influencing intestinal absorption and metabolism, such as the effect of metabolic enzymes (e.g. CYP450) and transport proteins (e.g. uptake and efflux), should be regarded to improve bioavailability predictions. Furthermore, the development of novel descriptors is associated with the structural variability of the dataset and the use of specific software (e. g. MODESLAB), where it is possible to identify substructural patterns that can better explain metabolic processes [4].

Finally, with respect to QSPR studies, the selection of the most appropriate modeling approach could play an important role in their predictability. This selection depends on the level of interpretability, accuracy, or applicability required. Although many QSPR models have been published to predict bioavailability, this property is a composite parameter that depends on

permeability, intestinal absorption, and first-pass metabolism, and it is difficult to identify the influence of each individual property from an overall bioavailability model. An appropriate variant might be to use individual models for each property, and then consider how the concerted action of all models can influence overall bioavailability. Nevertheless, Hou et al. [51] have suggested that this variant is really difficult when the prediction of some of the mechanisms involved in oral bioavailability is not reliable. A better option might be to develop different computational models to predict oral bioavailability and then propose a final classification using a consensus model [26].

At this point, it is important to highlight some aspects to be considered in the future development of QSPR and rule-based bioavailability models: i) the quality and size of the bioavailability database should be improved on the basis of data reported in reliable experimental assays; ii) variability between and within laboratories should be taken into account; iii) the active transport mechanism should be analyzed; iv) the combination of different *in silico* models for the same property should improve confidence in the final prediction of bioavailability; and v) *in silico* results should be assessed by experimental assays and subsequently used to increase the dataset and the accuracy of a future bioavailability model.

Like QSPR studies, PBPK models are very useful during the early drug discovery process and in the development of drug candidates. The PBPK model for predicting oral bioavailability prior to 'first in human' studies is very important for the decision-making process. Although, at later stages of drug development, combining *in silico* data with *in vitro* and *in vivo* measurements may provide better predictions of bioavailability, the final results depend on the extent, quality, and relevance of the input data [26]. In this regard, the integrity of the PK data, the input parameters, and the method of analysis must be well documented.

The currently available PBPK software must be selected correctly. An in-house program would be more suitable for use at the drug discovery stage, due to its flexibility for research. These programs make possible to develop their own models to explain experimental results or to propose new hypotheses on the absorption mechanism, etc. However, one of its drawbacks is related to programming codes, although programs such as MatLab™, Stella™, etc. offer more flexible programming capabilities [76].

Commercial programs such as SimCYP™, GastroPlus™, and PK-Sim™ are more suitable for the drug development stage. It is based on its user-friendly graphical interface, the ability to conduct virtual clinical trials, and its wide use by pharmaceutical companies to facilitate knowledge exchange [76].

The use of commercial software must be done with care due to some misconceptions related to the processes of solubility, permeability, and dissolution, which appear when considering the default functions and settings [76].

Accurate prediction of oral bioavailability, from chemical structure alone, remains a challenge in drug discovery today, because prediction of disposition parameters (e.g. clearance) is the most problematic input of the PBPK model. In this regard, the integration of consensus QSPR models to predict input properties with PBPK models can provide a more realistic simulation of the absorption and metabolism processes and is a valuable alternative to give mechanistic information on the oral bioavailability of drugs.

Funding

MA Cabrera-Perez is supported via project 10003 'Development of computational and experimental methodologies to demonstrate the therapeutic interchangability of multi-source pharmaceuticals on the Cuban List of Essential Medicines' funded by the Chemical Bioactive Center, Cuba. H Pham-The acknowledges principal financial support from the National Foundation for Science and Technology of Vietnam (NAFOSTED, Grant number 108.05-2017.13).

Declaration of Interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties. One referee declares that they are an employee of Pfizer Inc while another declares that they are an employee of Johnson & Johnson.

ORCID

Miguel Ángel Cabrera-Pérez http://orcid.org/0000-0001-5897-2230
Hai Pham-The http://orcid.org/0000-0003-4531-7223

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Waring MJ, Arrowsmith J, Leach AR, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. Nat Rev Drug Discov. Nature Publishing Group. 2015;14(7):475–486.
• **An update analysis of the factors reducing the high attrition rates in drug development**
2. Kubinyi H. Drug research: myths, hype and reality. Nat Rev Drug Discov. 2003;2(8):665–668.
3. Pritchard JF, Jurima-Romet M, Reimer ML, et al. Making better drugs: decision gates in non-clinical drug development. Nat Rev Drug Discov. 2003;2(7):542–553.
4. Cabrera-Pérez MA, Pham-The H, Bermejo M, et al. QSPR in oral bioavailability: specificity or integrality? Mini Rev Med Chem. 2012;12(6):534–550.
5. Zhu J, Wang J, Yu H, et al. Recent developments of in silico predictions of oral bioavailability. Comb Chem High Throughput Screen. 2011;14(5):362–374.
6. Committee for Medicinal Products for Human Use (CHMP). Guideline on the Investigation of Bioequivalence. 20 Jan 2010. London; European Medicines Agency (EMA). 2000.
7. Lin JH, Chiba M, Baillie TA. Is the role of the small intestine in first-pass metabolism overemphasized? Pharmacol Rev. 1999;51(2):135–158.
8. Chaturvedi PR, Decker CJ, Odinecs A. Prediction of pharmacokinetic properties using experimental approaches during early drug discovery. Curr Opin Chem Biol. 2001;5(4):452–463.
9. Hou T, Wang J, Zhang W, et al. ADME evaluation in drug discovery. 7. Prediction of oral absorption by correlation and classification. J Chem Inf Model. 2007;47(1):208–218.
10. Wang J, Hou T. Advances in computationally modeling human oral bioavailability. Adv Drug Deliv Rev. 2015;86:11–16.
•• **Excellent review about the development of in silico models to predict oral bioavailability.**

11. Agoram B, Woltosz WS, Bolger MB. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. Adv Drug Deliv Rev. 2001;50(Suppl 1):S41–S67.
12. Tropsha A. Best practices for QSAR model development, validation, and exploitation. Mol Inf. 2010;29(6–7):476–488.
•• **This review examines most critical QSAR modeling routines that should be considered as best practices in the field.**
13. Canault B, Bourg S, Vayer P, et al. Comprehensive network map of ADME-Tox databases. Mol Inform. 2017;36:1–8.
14. Hirono S, Nakagome I, Hirano H, et al. Non-congeneric structure-pharmacokinetic property correlation studies using fuzzy adaptive least-squares: oral bioavailability. Biol Pharm Bull. 1994;17(2):306–309.
15. Bertz RJ, Granneman GR. Use of in vitro and in vivo data to estimate the likelihood of metabolic pharmacokinetic interactions. Clin Pharmacokinet. 1997;32(3):210–258.
16. Pintore M, Van De Waterbeemd H, Piclin N, et al. Prediction of oral bioavailability by adaptive fuzzy partitioning. Eur J Med Chem. 2003;38(4):427–431.
17. Turner JV, Glass BD, Agatonovic-Kustrin S. Prediction of drug bioavailability based on molecular structure. Anal Chim Acta. 2003;485:89–102.
18. Turner JV, Maddalena DJ, Agatonovic-Kustrin S. Bioavailability prediction based on molecular structure for a diverse series of drugs. Pharm Res. 2004;21(1):68–82.
19. Wang J, Krudy G, Xie XQ, et al. Genetic algorithm-optimized QSPR models for bioavailability, protein binding, and urinary excretion. J Chem Inf Model. 2006;46(6):2674–2683.
20. Moda TL, Montanari CA, Andricopulo AD. Hologram QSAR model for the prediction of human oral bioavailability. Bioorg Med Chem. 2007;15(24):7738–7745.
21. Ma CY, Yang SY, Zhang H, et al. Prediction models of human plasma protein binding rate and oral bioavailability derived by using GA-CG-SVM method. J Pharm Biomed Anal. 2008;47(4–5):677–682.
22. Xu X, Zhang W, Huang C, et al. A novel chemometric method for the prediction of human oral bioavailability. Int J Mol Sci. 2012;13(6):6964–6982.
23. Olivares-Morales A, Hatley OJD, Turner D, et al. The use of roc analysis for the qualitative prediction of human oral bioavailability from animal data. Pharm Res. 2014;31(3):720–730.
24. Musther H, Olivares-Morales AA, Hatley OJD, et al. Animal versus human oral drug bioavailability: do they correlate? Eur J Pharm Sci. 2013;57(100):280–291.
25. Olah M, Halip L, Hădărugă N, et al. Chemical Informatics: WOMBAT and WOMBAT-PK: Bioactivity Databases for Lead and Drug Discovery. In: Schreiber SL, Kapoor T, Wess G, editors. Chemical Biology: From Small Molecules to Systems Biology and Drug Design. Weinheim, Germany: Wiley VCH Verlag GmbH & Co. 2007:760–786.
26. Cabrera-Pérez MÁ, González-Álvarez I. Computational and pharmacoinformatic approaches to oral bioavailability prediction. In: Hu M, Li X, editors. Oral bioavailability: basic principles, advanced concepts, and applications. Hoboken, New Jersey: John Wiley & Sons, Inc.; 2011: 519–534.
27. Hou T, Wang J, Zhang W, et al. 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.
28. Tian S, Li Y, Wang J, et al. ADME evaluation in drug discovery. 9. Prediction of oral bioavailability in humans based on molecular properties and structural fingerprints. Mol Pharm. 2011;8(3):841–851.
29. Sietsema WK. The absolute oral bioavailability of selected drugs. Int J Clin Pharmacol Ther Toxicol. 1989;27(4):179–211.
• **One of the most reliable databases used to develop computational models of oral bioavailability.**
30. Zhao YH, Le J, Abraham MH, et al. 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.
31. Moda TL, Torres LG, Carrara AE, et al. PK/DB: database for pharmacokinetic properties and predictive in silico ADME models. Bioinformatics. 2008;24(19):2270–2271.
32. Cao D, Wang J, Zhou R, et al. ADMET evaluation in drug discovery. 11. Pharmacokinetics knowledge base (PKKB): a comprehensive database of pharmacokinetic and toxic properties for drugs. J Chem Inf Model. 2012;52:1132–1137.
33. Todeschini R, Consonni V. Handbook of molecular descriptors. Weinheim, Germany: Wiley VCH; 2000.
34. Norinder U, Bergstrom CA. Prediction of ADMET properties. ChemMedChem. 2006;1(9):920–937.
35. OECD. Guidance document on the validation of (quantitative) structure-activity relationships [(Q)SAR] models. OECD series on testing and assessment no. 69. ENV/JM/MONO(2007)2. Paris, France: Organisation for Economic Cooperation and Development; 2007.
36. Veber DF, Johnson SR, Cheng HY, et al. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem. 2002;45(12):2615–2623.
37. Martin YC. A bioavailability score. J Med Chem. 2005;48(9):3164–3170.
38. Varma MV, Obach RS, Rotter C, et al. Physicochemical space for optimum oral bioavailability: contribution of human intestinal absorption and first-pass elimination. J Med Chem. 2010;53(3):1098–1108.
39. Kier LB, Hall LH. An electrotopological-state index for atoms in molecules. Pharm Res. 1990;7(8):801–807.
40. Estrada E. How the parts organize in the whole? A top-down view of molecular descriptors and properties for QSAR and drug design. Mini Rev Med Chem. 2008;8(3):213–221.
41. Tao L, Zhang P, Qin C, et al. Recent progresses in the exploration of machine learning methods as in-silico ADME prediction tools. Adv Drug Deliv Rev. 2015;86:83–100.
42. Ursu O, Rayan A, Goldblum A, et al. Understanding drug-likeness. WIREs Comput Mol Sci. 2011;1:760–781.
43. Doak BCC, Over BB, Giordanetto F, et al. Oral druggable space beyond the rule of 5: insights from drugs and clinical candidates. Chem Biol. 2014;21(9):1115–1142.
44. Giordanetto F, Kihlberg J. Macrocyclic drugs and clinical candidates what can medicinal chemists learn from their properties. J Med Chem. 2014;57(2):278–295.
45. Pham-The H, González-Álvarez I, Bermejo M, et al. The use of rule-based and QSPR approaches in adme profiling: a case study on caco-2 permeability. Mol Inform. 2013;32(5–6):459–479.
46. Lu JJ, Crimin K, Goodwin JT, et al. Influence of molecular flexibility and polar surface area metrics on oral bioavailability in the rat. J Med Chem. 2004;47:6104–6107.
47. Yoshida F, Topliss JG. QSAR model for drug human oral bioavailability. J Med Chem. 2000;43(13):2575–2585.
48. Andrews CW, Bennett L, Yu LX. Predicting human oral bioavailability of a compound: development of a novel quantitative structure-bioavailability relationship. Pharm Res. 2000;17(6):639–644.
49. Wang Z, Yan A, Yuan Q, et al. Explorations into modeling human oral bioavailability. Eur J Med Chem. 2008;43(11):2442–2452.
50. Imawaka H, Ito K, Kitamura Y, et al. Prediction of human bioavailability from human oral administration data and animal pharmacokinetic data without data from intravenous administration of drugs in humans. Pharm Res. 2009;26(8):1881–1889.
51. Hou T, Wang J. Structure-ADME relationship: still a long way to go? Expert Opin Drug Metab Toxicol. 2008;4(6):759–770.
52. Peters SA. Physiologically-based pharmacokinetic (PBPK) modeling and simulations. principles, methods, and applications in the pharmaceutical industry. New Jersey: John Wiley & Sons, Inc; 2012.
53. Bouzom F, Ball K, Perdaems N, et al. Physiologically based pharmacokinetic (PBPK) modelling tools: how to fit with our needs? Biopharm Drug Dispos. 2012;33:55–71.
• **Good paper to show the relevance of PBPK models.**
54. Lin L, Wong H. Predicting oral drug absorption: mini review on physiologically-based pharmacokinetic models. Pharmaceutics. 2017;9(41):1–14.

55. Espié P, Tytgat D, Sargentini-Maier M-L, et al. Physiologically based pharmacokinetics (PBPK). Drug Metab Rev. 2009;41(3):391–407.
•• **Excellent review about PBPK models.**
56. Theil FP, Guentert TW, Haddad S, et al. Utility of physiologically based pharmacokinetic models to drug development and rational drug discovery candidate selection. Toxicol Lett. 2003;138(1–2):29–49.
57. Peters SA. Evaluation of a generic physiologically based pharmacokinetic model for lineshape analysis. Clin Pharmacokinet. 2008;47(4):261–275.
58. Jones HM, Chen Y, Gibson C, et al. Physiologically based pharmacokinetic modeling in drug discovery and development: a pharmaceutical industry perspective. Clin Pharmacol Ther. 2015;97(3):247–262.
59. Kesisoglou F, Chung J, Asperen JV, et al. Physiologically based absorption modeling to impact biopharmaceutics and formulation strategies in drug development d industry case studies. J Pharm Sci. 2016;105(9):2723–2734.
60. Huang SM, Abernethy DR, Wang Y, et al. The utility of modeling and simulation in drug development and regulatory review. J Pharm Sci. 2013;102(9):2912–2923.
61. Dokoumetzidis A, Valsami G, Macheras P. Modelling and simulation in drug absorption processes. Xenobiotica. 2007;37:1052–1065.
62. Yu LX, Amidon GL. A compartmental absorption and transit model for estimating oral drug absorption. Int J Pharm. 1999;186(2):119–125.
63. Yu LX. An integrated model for determining causes of poor oral drug absorption. Pharm Res. 1999;16(12):1883–1887.
64. Parrott N, Paquereau N, Coassolo P, et al. An evaluation of the utility of physiologically based models of pharmacokinetics in early drug discovery. J Pharm Sci. 2005;94(10):2327–2343.
65. Paixao P, Gouveia LF, Morais JAG. Prediction of the human oral bioavailability by using in vitro and in silico drug related parameters in a physiologically based absorption model. Int J Pharm. 2012;429:84–98.
66. Thelen K, Coboeken K, Willmann S, et al. Evolution of a detailed physiological model to simulate the gastrointestinal transit and absorption process in humans, part 1: oral solutions. J Pharm Sci. 2011;100(12):5324–5345.
67. Thelen K, Coboeken K, Willmann S, et al. Evolution of a detailed physiological model to simulate the gastrointestinal transit and absorption process in humans, part ii: extension to describe performance of solid dosage forms. J Pharm Sci. 2012;101(3):1267–1280.
68. Jamei M, Turner D, Yang J, et al. Population-based mechanistic prediction of oral drug absorption. Aaps J. 2009;11(2):225–237.
69. Darwich AS, Neuhoff S, Jamei M, et al. Interplay of metabolism and transport in determining oral drug absorption and gut wall metabolism: a simulation assessment using the "Advanced Dissolution, Absorption, Metabolism (ADAM)" Model. Curr Drug Deliv. 2010;11:716–729.
70. Cai H, Stoner C, Reddy A, et al. Evaluation of an integrated in vitro-in silico PBPK (physiologically based pharmacokinetic) model to provide estimates of human bioavailability. Int J Pharm. 2006;308(1–2):133–139.
71. Daga PR, Bolger MB, Haworth IS, et al. Physiologically-based pharmacokinetic modeling in lead optimization I: evaluation and adaptation of gastroplus to predict bioavailability of medchem series. Mol Pharm. 2018;15:821–830.
72. Olivares-Morales A, Ghosh A, Aarons L, et al. Development of a novel simplified PBPK absorption model to explain the higher relative bioavailability of the oros ® formulation of oxybutynin. Aaps J. 2016;18(6):1532–1549.
73. Parrott N, Grimm HP, Fleury A, et al. Challenges and opportunities with modelling and simulation in drug discovery and drug development. Xenobiotica. 2007;37:1295–1310.
74. Ahmed SSSJ, Ramakrishna V. Systems biological approach of molecular descriptors connectivity: optimal descriptors for oral bioavailability prediction. PLoS One. 2012;7(7):1–10.
75. Kim MT, Sedykh A, Chakravarti SK, et al. Critical evaluation of human oral bioavailability for pharmaceutical drugs by using various cheminformatics approaches. Pharm Res. 2014;31:1002–1014.
76. Sugano K. Introduction to computational oral absorption simulation. Expert Opin Drug Metab Toxicol. 2009;5(3):259–293.