Advanced Drug Delivery Reviews xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Advanced Drug Delivery Reviews

journal homepage: www.elsevier.com/locate/addr

Advances in computationally modeling human oral bioavailability ☆

Junmei Wang [a,*], Tingjun Hou [b,c]

[a] Department of Biochemistry, The University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd. Dallas, TX 75390, USA
[b] Institute of Functional Nano and Soft Materials (FUNSOM), Jiangsu Key Laboratory for Carbon-Based Functional Materials and Devices and Collaborative Innovation Center of Suzhou Nano Science and Technology, Soochow University, Suzhou, Jiangsu 215123, China
[c] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China

ARTICLE INFO

Available online xxxx

Keywords:
Human oral bioavailability (HOBA)
Human intestinal absorption (HIA)
ADME–Tox
QSAR
In silico modeling computer-aided drug design

ABSTRACT

Although significant progress has been made in experimental high throughput screening (HTS) of ADME (absorption, distribution, metabolism, excretion) and pharmacokinetic properties, the ADME and Toxicity (ADME–Tox) in silico modeling is still indispensable in drug discovery as it can guide us to wisely select drug candidates prior to expensive ADME screenings and clinical trials. Compared to other ADME–Tox properties, human oral bioavailability (HOBA) is particularly important but extremely difficult to predict. In this paper, the advances in human oral bioavailability modeling will be reviewed. Moreover, our deep insight on how to construct more accurate and reliable HOBA QSAR and classification models will also discussed.

Contents

1. Introduction

It is estimated that the entire chemical space exceeds 10^{60} molecules, and it is impossible to synthesize all of them given the fact that the total weight of earth is only about 6.0×10^{27} g. As a matter of fact, only 27 million compounds have been registered [1]. Even though the synthesized compounds only occupy a tiny fraction of the entire chemical space, it is much larger than the biological chemical space due to the fact that there are a few thousands of small molecules within our own bodies. As the biological chemical space only represent an amazingly small fraction of the entire chemical space, it is understandable that to discover small molecules that efficiently interact with protein targets is a very difficult task. Although numerous

Abbreviations: *ADME–Tox*, Absorption, distribution, metabolism, excretion, and toxicity; *HOBA*, Human oral bioavailability; *HIA*, Human intestinal absorption; QSAR, Quantitative structure–activity relationship; *MLR*, Multiple linear regressions; *GA*, Genetic algorithm; *LOO*, Leave-one-out; R^2, Regression coefficient; Q^2, Cross-validation regression coefficient; *AUE*, Average unsigned error; *RMSE*, Root-mean-square error; AUE_{test}, Average unsigned error of the test set; $RMSE_{test}$, Root-mean-square error of the test set; *N*, Number of data points; N_{test}, Number of data points in a test set; *LogP*, Logarithm of octanol–water partition coefficient; *LogD*, Logarithm of water distribution coefficient

☆ This review is part of the *Advanced Drug Delivery Reviews* theme issue on "In silico ADMET predictions in pharmaceutical research".

* Corresponding author. Tel.: +1 214 6484146.
E-mail address: junmei.wang@utsouthwestern.edu (J. Wang).

http://dx.doi.org/10.1016/j.addr.2015.01.001
0169-409X/

new technologies, such as combinatorial chemistry, high throughput screening and computer-aided drug design have been applied to facilitate the discovery of new drugs, the number of new molecular entities approved annually by FDA (U.S. Food and Drug Administration) has not changed significantly in the last two decades. What are the major reasons that cause the attrition of drug candidates during clinical trials? The lack of efficiency and poor ADME–Tox (absorption, distribution, metabolism, excretion, and toxicity) and pharmacokinetics are responsible for most of the drug attrition [2].

Among the many ADME–Tox/PK properties, bioavailability is particularly important for the orally administered drugs. Today, high throughput screenings of human oral bioavailability (HOBA) are routinely conducted in pharmaceutical companies. However, the in vitro and in vivo assays are much time consuming and costly. Only a tiny fraction of synthesized and screened compounds are selected to do the analysis. In silico HOBA modeling, on the other hand, is much more efficient and can deal with large screening libraries. Moreover, in silico HOBA models can serve as drug likeness filters to prioritize screening libraries. Those filters typically have better discriminative power than the conventionally used drug likeness filters, like Lipinski's 'Rule of Five' [3]. It is a trend that in silico ADME–Tox models, particularly HOBA, are incorporated into the paradigm of drug lead identification and optimization procedures [4–16].

1.1. ADME–Tox

As one of the hot fields in computer-aided drug design (CADD), numerous reviews have been published recently on the progress of ADME–Tox modeling [17–21], here in this paper we only focus on the latest advances of in silico modeling of human oral bioavailability. ADME–Tox properties can be broadly classified into two categories, namely, the "physicochemical" and "physiological". The physicochemical properties, including aqueous solubility, logarithm of octanol–water partition coefficient (*logP*), logarithm of octanol–water distribution coefficient (*logD*), pKa, etc., are governed by simple physicochemical laws. The physiological ADME–Tox properties can be further grouped into in vitro ADME–Tox properties (such as Caco-2 permeability and MDCK permeability, liver microsomes) and in vivo pharmacokinetic properties (such as oral bioavailability, human intestinal absorption—HIA, plasma protein binding—PPB, urinary excretion, area under the plasma concentration–time curve (AUC), total body clearance (Cl), volume of distribution, elimination half time ($t_{1/2}$)). As physiological ADME–Tox properties, particularly oral bioavailability, are governed by many factors, it is a very challenging task to adequately model and accurately predict the physiological ADME–Tox properties.

1.2. Human oral bioavailability

Oral bioavailability (OBA) is one of the most important pharmacokinetic properties in drug discovery. As the oral form is the most convenient way to administrate a drug, it is not a surprise that about 80% of the dosage forms in the worldwide market are administrated orally [22]. OBA represents the percentage of an oral dose that is available to produce pharmacological actions. In practice, OBA is defined as the fraction of the oral dose that reaches the system circulation in an active form and measured by the ratio of the dose-corrected AUC (area under curve) of the oral route to that of the intravenous route. For an oral drug, the amount of the active form that reaches the system circulation is reduced due to incomplete absorption in gastrointestinal track and the first-pass metabolism. Therefore, oral bioavailability is ranged from 0 to 100%.

It is a very challenging task to adequately model and accurately predict the oral bioavailability of a drug because this physiological property is a complex function of many biological and physicochemical properties, which include the aqueous solubility of the drug in the gastrointestinal tract, the intestinal membrane permeability, and the extent of the first-pass metabolism which occurred in the liver, gut and intestine, and even the dosage form of the drug. Moreover, the measurement of the oral bioavailability of a drug is affected by other factors like whether the drug is taken with or without food, whether other drugs are taken concurrently, as well as the disease states. Those factors may alter the drug absorption, and the liver metabolism. For example, the oral bioavailability of patients with liver disease may be increased due to the reduced liver metabolism. The above factors may vary from patient to patient and from time to time for the same patient. This complicate picture explains why the measurement errors of oral bioavailability are very large. According to the survey of 367 drugs conducted by Wang et al.- [23], the average unsigned error and root-mean square error of the experimental measurements are 12.1 and 14.5%, respectively.

In order to develop an oral drug with high bioavailability, medicinal chemists apply a simple rule to select drug candidates: those having high aqueous solubility and high membrane permeability tend to have high OBA; those having low aqueous solubility and low permeability tend to have poor OBA; and the others might need careful formulation to improve their dissolution or absorption rate. This simple rule is based on the fact that drug dissolution and permeability control the rate and extent of drug absorption in the GI track. Certainly, a drug with high oral bioavailability should also be largely free from fast-pass metabolism.

1.3. In silico models of HOBA prediction

Attempts have been made to predict HOBA back to year 2000 by Andrews, Bennett and Xu [24], and Yoshida and Topliss [25,26]. Later on, numerous models were published [25,27,28] and reviewed by ourselves [29], and others [30]. The following is a brief summary of HOBA models developed prior to 2008: most models were developed using relatively small data sets ($n < 600$) and they merely make reliable prediction for the compounds in the screening libraries. For the classification models developed before 2008, the rates of the correct assignment are usually lower than 70%; for the QSAR models, the RMSEs are ranged from 24 to 30%. In the following sections, we will present reviews on the latest HOBA models.

2. Recent advances in HOBA modeling

In 2008, a classifier was developed by Ma et al. with GA (genetic algorithm)–CG (conjugated gradient)–SVM (support vector machine) method for 866 compounds that have human oral bioavailability data [31]. GA was applied to select descriptors that were calculated using Cerius 2 software package (http://www.accelrys.com), while SVM was used to construct classification model and CG was applied to optimize the parameters of kernel functions of SVM. The prediction accuracy, 80% for the training set (690 compounds) and 86% for the test set (76 compounds) is encouraging. However, the classifier has poor performance for the "negative" class: the prediction accuracy is only 44% and the false positive (FP) is even larger than true negative (TN). This phenomenon can be explained by that fact that a very small cutoff of 20% was applied to assign 'positive' and 'negative' classes. Even if the prediction accuracy is good, it cannot be used to further discriminate the compounds that belong to the "positive" class.

In 2009, a set of predictive models for human bioavailability were developed by Imawaka et al. using the human oral administration data and animal pharmacokinetic data as descriptors [32].

$$OBA = \frac{AUC_{po}/Dose_{po}}{AUC_{iv}/Dose_{iv}} = \frac{CL_{tot}}{CL_{po}} = \frac{\beta \times Vd_{\beta}}{CL_{po}} \quad (1)$$

Please cite this article as: J. Wang, T. Hou, Advances in computationally modeling human oral bioavailability, Adv. Drug Deliv. Rev. (2015), http://dx.doi.org/10.1016/j.addr.2015.01.001