informa
healthcare

Material Code: 5637

Expert Opinion on Drug Metabolism & Toxicology colour figure order form

All figures will appear in colour online free of charge. Print issue colour reproduction will only be available if the author contributes to the associated costs. Please complete this form if you require any figures in your manuscript to be printed in colour. Completed forms should be returned to the Production Editor.

Article title:

Manuscript ID: 313889

To be published in: EOMT Issue.....6..... Volume.....4.....

I require the following number of figures to be printed in colour:____
Figues to be printed in colour are figure numbers:_____
Please charge: £55 / $108 / €90 per figure

Colour Figures	Total Cost

Invoice Delivery Address (including email and phone number):

Credit Card Payment:

Credit Card ☐ Please charge my AMEX/VISA/MasterCard (delete as appropriate).

Card No. ☐☐☐☐/☐☐☐☐/☐☐☐☐/☐☐☐☐ Exp.date ☐☐/☐☐

CVV number (last 3 or 4 digits on reverse of card):
Name on credit card:
Address where registered:
Signature:

Date:

All orders are accepted subject to Informa Healthcare's standard Terms and Conditions.

Please fax back to: **Adam Hooper**
Fax no +44 20 7017 7667

Journal: Expert Opinion on Drug Metabolism & Toxicology (EOMT)

Title: Structure – ADME relationship: still a long way to go?

ID: 313889

Corresponding author: Tingjun Hou

AUTHOR: Please complete the colour figure order form that accompanies your proof if you require figures to appear in colour in the print version of the journal; once you have completed your article and credit card details, email or fax the form back to the production editor. The following queries have arisen during the editing of your manuscript. Please answer the queries by making the necessary corrections on the CATS online corrections form. Once you have added all your corrections, please press the SUBMIT button.

Page number	Line number	Query
1	25	Are you happy with the slight rewording here?
10	Ref 17	Any update on publication?
5	342	Check the symbol is ok.
		References [26-69] have been renumbered to meet the requirement for numbered references that first citations be in numerical order. Please check that each citation number points to the correct reference.

Expert Opinion

1. Background
2. Prediction of solubility
3. Prediction of human intestinal absorption
4. Prediction of oral bioavailability
5. Expert opinion

Structure – ADME relationship: still a long way to go?

Tingjun Hou† & Junmei Wang
†*University of California at San Diego, Department of Chemistry and Biochemistry, Center for Theoretical Biological Physics, 9500 Gilman Drive, La Jolla, CA, USA*

Background: Theoretical models for predicting absorption, distribution, metabolism and excretion (ADME) properties play increasingly important roles in support of the drug development process. *Objective*: We briefly review the *in silico* prediction models for three important ADME properties, namely, aqueous solubility, human intestinal absorption, and oral bioavailability. *Methods*: Rather than giving detailed descriptions of the ADME prediction models, we focus on the discussions of the prediction accuracies of the *in silico* models. *Results/conclusion*: We find that the robustness and predictive capability of the ADME models are directly associated with the complexity of the ADME property. For the ADME properties involving complex phenomena, such as bioavailability, the *in silico* models usually cannot give satisfactory predictions. Moreover, the lack of large and high-quality data sets also greatly hinder the reliability of ADME predictions. While considerable progress has been achieved in ADME predictions, many challenges remain to be overcome.

Keywords: ADME, bioavailability, intestinal absorption, QSPR, solubility

Expert Opin. Drug Metab. Toxicol. (2008) 4(6):1-12

1. Background

The importance of optimizing the absorption, distribution, metabolism and excretion (ADME) properties for potential drug candidates have been widely recognized [1]. The success of a drug is determined not only by good efficacy and specificity, but also by having acceptable ADME and toxicity properties (ADMET) [2]. Traditionally, in the drug discovery process, the efficacy and specificity of a drug candidate are usually evaluated at the early stage, then the ADMET properties are considered at a relatively late stage [3]. The traditional 'serial' diagram of drug discovery usually results in a high rate of attrition in the later stages of drug discovery, where the costs increase dramatically. Analysis of the failure of new chemical entities (NCEs) shows that the poor ADMET properties are the major cause of failure of new pharmacologically promising compounds [4].

The traditional diagram began to change in the early 1990s. Pharmaceutical scientists try to optimize the ADMET properties early in the drug discovery process, as well as efficacy and specificity as a 'parallel' diagram. A recent analysis shows that the attrition rate caused by the adverse pharmacokinetic and bioavailability aspects has been greatly decreased [5]. Overall, the failure of developing candidates due to improper ADME/formulation, toxicology, and safety issues decreased from approximately 60% in 1991 to around 45% in 2000 [5].

Over the last 10 years, a number of high-throughput (HT) experimental techniques have been developed to evaluate the ADME properties, such as the Caco-2 permeability screening based on the 3-day Caco-2 culture system [6], high-throughput kinetic solubility assay [7,8], metabolic stability screening using microsomes or hepatocytes [9], and liquid chromatography-mass spectroscopy

informa healthcare

(LCMS) and fluorogenic assays through cytochrome cYP inhibition for metabolism related to drug–drug interactions [10,11]. Although much progress has been made in HT ADME experimental assays, compared with high-throughput screening (HTS) activity assays or combinatorial synthesis, the ADME experiments still have low throughput capacity, thus limiting the application of these assays to only a fraction of compounds in drug discovery. Therefore, *in silico* prediction models for the ADME properties are urgently needed to alleviate the bottlenecks in ADME experiments. *In silico* models have great potential to predict *in vitro* and *in vivo* ADME properties quickly to assist in prioritizing the large numbers of compounds, and no experiments are necessary.

The ADME predictions at the early stage usually focus on some simple ADME or ADME-related properties, such as octanol-water partitioning coefficient (log*P*), apparent partition coefficient (log*D*), intrinsic solubility (log*S*), etc. As a result of the increase in available experimental data in the literature, considerable efforts have been made to predict more 'complex' ADME properties, such as human intestinal absorption, blood–brain partitioning, oral bioavailability, clearance, volume of distribution, and metabolism. The ADME properties for which the prediction models have been developed are shown in **Figure 1**. In recent years, an increasing number of *in silico* models for predicting ADME properties have been reported [12], including solubility [13-15], Caco-2 permeability [15,16], human intestinal absorption [14-18], oral bioavailability [14,17], blood–brain partitioning [16,19], *P*-glycoprotein-mediated transport [14,20], plasma-protein binding [12], metabolism [21], volume of distribution [12], clearance [12], and even half-life [12]. Meanwhile, a number of computational software systems that can predict a range of ADME properties have been released (see **Table 1**).

Quantitative structure–property relationship (QSPR) approaches have been widely applied for modeling most ADME properties. Three essential components responsible for the quality of an ADME prediction model – namely, the data set to be used to generate the model; the descriptors of molecular structures to be used to characterize the properties of the molecule and to be correlated with the experimental data; and the statistical techniques to generate the model – have been discussed extensively [15,16,22]. The QSPR approaches simply construct the relationships between the molecular structures and the ADME properties, and do not necessarily know the underlying mechanism of an ADMET property. In contrast, molecular modeling approaches have been used to investigate the possible underlying molecular mechanism of a specific property or to understand the potential interactions between the small molecules and proteins involved in ADME processes by using molecular mechanics, molecular dynamics, pharmacophore modeling, molecular docking, homology modeling, or even quantum mechanics calculations. So far molecular modeling approaches in ADME prediction have only been applied on very limited ADME properties, especially metabolism related to cytochrome P450s.

Since a lot of *in silico* models for ADME predictions are available, two questions may be raised: what are the prediction accuracies of these models, and can they be effectively used in the pharmaceutical industry? In this review, the *in silico* prediction models for three representative ADME properties are reviewed, which include solubility, human intestinal absorption (HIA), and oral bioavailability. Rather than giving detailed descriptions of the ADME prediction models, we focus mainly on the prediction accuracies of the *in silico* models. It should be noted that the ADME prediction models discussed here were developed using the traditional QSPR approaches. In the recent past, physiologically based pharmacokinetic (PBPK) models have received a lot of attention because they may give us valuable information on how the various factors influence PK [23]. PBPK models were not discussed here because this group of models usually needs extra physiological parameters from experiments and cannot be developed solely from molecular structures.

2. Prediction of solubility

The solubility of organic molecules in water has a significant impact on many ADME-related properties of drugs, such as absorption, distribution, transport and eventually bioavailability [24]. The solubility of a neutral compound or of a compound in its non-ionized form is defined as the intrinsic solubility and normally represented as log*S*, where *S* is the concentration of the compound in mol/l in a saturated aqueous solution in equilibrium with the most stable form of the crystalline material. In practice, about 85% of drugs have log*S* between -1 and -5, and virtually none has a value < -6. Empirically, the log*S* range of -1 to -5 for most drugs reflects a compromise between the polarity necessary for reasonable aqueous solubility and the hydrophobicity necessary for acceptable membrane transport [25].

Solubility and intestinal absorption are two counterparts applied in the Biopharmaceutics Classification System (BCS) [26]. According to the BCS, drug substances are classified as follows:

- Class I: High permeability, high solubility. These compounds are well absorbed, and their absorption secretion rate is usually higher than excretion.
- Class II: High permeability, low solubility. The bioavailability of these compounds is limited by their solubility.
- Class III: Low permeability, high solubility. The absorption is limited by the permeation rate, but the drug is solvated very fast.
- Class IV: Low permeability, low solubility. These compounds are usually not well absorbed over the intestinal mucosa, and a high variability is expected.

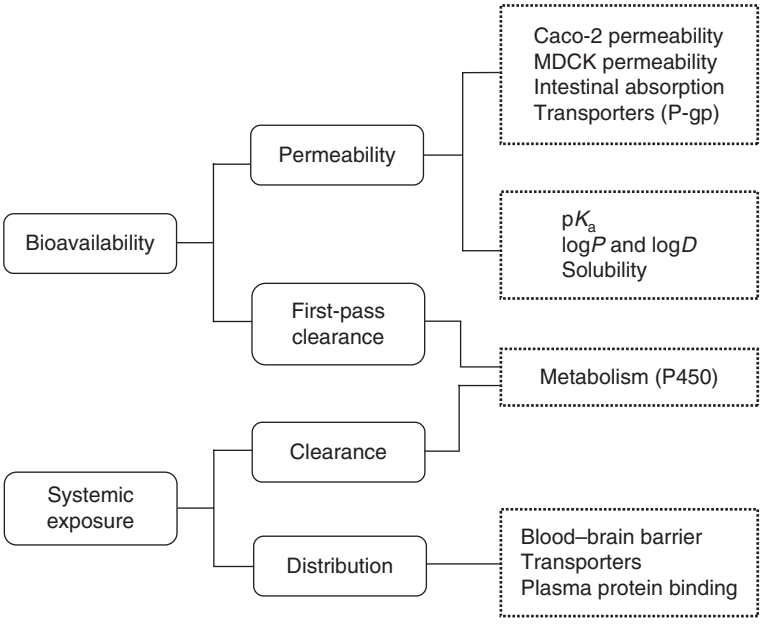

Figure 1. *In silico* prediction models for ADME properties. The figure does not give a logical flow of the ADME studies, but attempts to group them roughly into different prediction models.

Table 1. Popular commercial software available for predicting ADME and ADME-related properties.

Software	Company	pK_a	log*P*	log*D*	Sol	HIA	C2	BBB	Bio	Mtb	Trp	PPB	Others
Cerius2	Accelrys				√	√		√		√		√	√
Qikprop	Schrödinger		√		√		√	√				√	√
ACD/labs	ACD Labs	√	√	√	√								
Volsurf	Tripos				√	√	√	√		√		√	
ADME Boxes	PharmaAlgorithms	√	√	√	√	√	√		√		√	√	
ADME Predictor	Simulations Plus	√	√	√	√	√		√				√	√
KnowItAll	Bio-Rad Lab	√	√	√	√	√		√	√			√	√
NorayMet ADME	Noraybio	√	√	√	√	√	√			√		√	√
iDEA ADME	LION Bioscience				√	√	√	√	√	√			
PreADMET			√		√	√	√	√				√	√
ADME Collection	Scitegic				√	√		√		√		√	
Jchem	ChemAxon	√	√	√									
StarDrop	Biofocus	√	√	√	√	√		√		√	√	√	√
ADMEWORKS	Fujitsu		√		√	√		√		√			√
Meta	Multicase									√			
MetaSite	Molecular Discovery									√			

BBB: Blood–brain barrier permeability; Bio: Oral bioavailability; C2: Caco-2 permeability; Mtb: Metabolism; PPB: Plasma–protein binding; Sol: Solubility; Trp: Carrier-mediated transport.

It is worth mentioning that the FDA BCS guidance, based on the current definition of solubility, is not reliable enough for all drugs, especially many true BCS class I drugs [27,28].

Until now, a lot of models have been proposed for the prediction of solubility. These models can be roughly divided into three categories: experiment-related models, QSPR-based models, and atom/group contribution models.

2.1 Experiment-related models

The models in the first category calculate aqueous solubility using one or several experimental physicochemical properties such as partition coefficient, melting points, boiling points, or molar volumes; for example, the general solubility equation (GSE) proposed by Yalkowsky and co-workers [29,30]. GSE related the molar intrinsic solubility (S_w) to the Celsius melting point (*mp*) and the octanol partition coefficient (log*P*) by the following simple equation:

$$\log S_w = 0.5 - 0.01(mp - 25) - \log P \quad (1)$$

GSE has been shown to produce reasonable predictions for a wide variety of compounds. These methods require the experimental values for *mp*, so they are not applicable to compounds not yet synthesized or isolated. Therefore, these methods only have limited application domains.

2.2 QSPR-based models

The second category of models tries to construct the prediction models by correlating solubility with a diverse set of descriptors such as physiochemical descriptors and molecular properties by various statistical techniques. The methods developed by Huuskonen *et al.* [31], Abraham *et al.* [32], Jorgensen *et al.* [33], McElroy *et al.* [34], McFarland *et al.* [35], Liu *et al.* [36], Tetko *et al.* [37], Engkvist *et al.* [38], Yan *et al.* [39], Butina *et al.* [40], Goller *et al.* [41], among others, belong to this class.

Equation 2 is the solubility prediction model developed by Abraham and Le [32]. The model can give good prediction for 65 compounds in the test set, indicated by a standard deviation (*SD*) value of 0.496:

$$\log S = 0.510 - 1.020R_2 + 0.813\pi_2^H + 2.124\sum\alpha_2^H + 4.187\sum\beta_2^H - 3.337\sum\alpha_2^H \times \sum\beta_2^H - 3.986V_x \quad (2)$$
$$(n = 594,\ SD = 0.562,\ r^2 = 0.918,\ F = 1089,\ AAE = 0.409)$$

In Equation 2, π_2^H is the dipolarity/polarizability; $\sum\alpha_2^H$ is the overall or summation hydrogen bond acidity; $\sum\beta_2^H$ is the overall or summation hydrogen bond basicity; V_x is the McGowan characteristic volume; *SD* is the standard deviation; r^2 is the squared correlation coefficient; *F* is *F*-value of the Fisher's exact test; *AAE* is the average absolute error. The term $\sum\alpha_2^H \times \sum\beta_2^H$ was introduced to deal with hydrogen bond interactions between acidic and basic sites in the solid or liquid.

Another example of solubility prediction model developed by Jorgensen and co-workers is shown in Equation 3 [33]:

$$\log S = 0.3158\text{ESXL} + 0.6498\text{HBAC} + 2.192\#\text{amine} - 1.759\#\text{nitro} - 161.6\text{HBAC.HBDN}^{1/2}/\text{SASA} + 1.181 \quad (3)$$
$$(n = 1.50,\ r^2 = 0.88,\ q^2 = 0.87,\ rmse = 0.72)$$

In Equation 3, ESXL is the averaged solute-water Lennard-Jones (ESXL) interaction energies; HBDN and HBAC are the numbers of H-bond donors and acceptors of solute, respectively; #amine is the number of non-conjugated amine groups; #nitro is the number of nitro groups; SASA is the solvent-accessible surface area; q^2 is the predictive squared correlation coefficient based on leave-one-out cross-validation; *rmse* is the root mean square error.

The solubility prediction model developed by Huuskonen needs to be emphasized here because the data set reported by Huuskonen [42] has been widely used by other researchers to develop solubility prediction models. The 'Huuskonen' data set includes 1297 organic compounds extracted from the AQUASOL database and SCR's PHYSPROP database [42]. Huuskonen divided the whole data set into a training set of 884 compounds and a randomly chosen test set of 413 compounds. Molecular connectivity, shape, and atom-type electrotopological state (E-state) indices were used as structural parameters. A 30 – 12 – 1 artificial neural network using 24 atom-type E-state indices and six other topological indices gave the best performance, and better than the model using the multiple linear regression. The model can give a predictive r^2 = 0.92 and *SD* = 0.60 for a test set of 413 compounds and a *SD* = 0.63 for a 21-compound test set. Please note that the test set of 413 compounds was used for controlling the training process of artificial neural network, and thus it is not the 'real' test set. So the Huuskonen's model was only validated by the 21-compound test set. Obviously, 21 is a limited number and a larger external test set is necessary to give more extensive evaluation of the predictive capability of the model.

The practical superiority of this type of method is that it does not require the knowledge of any experimental data of compounds, because all descriptors needed are calculated directly from a two-dimensional (2-D) or three-dimensional (3-D) molecular structure. However, this class of methods has its inherent deficiencies. First, the methods usually require many molecular descriptors, which may be difficult to obtain or can only be calculated by using a commercial software. For example, in the work of Engkvist *et al.*, the authors used a total of 63 physicochemical and topological descriptors [38]. The dependence of the descriptors calculated from other theoretical models poses some problems for estimating the solubility of a molecule using the models in the public domain and developing a program or scripts to

estimate solubility automatically. Secondly, the prediction accuracy of the QSAR model is closely related to the accuracy of descriptors used in the model. In addition, the relationship between the descriptors and the aqueous solubility is usually not straightforward.

2.3 Atom/group contribution models

The third class of models for predicting aqueous solubility is based on atom or group contribution. In principle, the atom/group contribution models are QSPR-based models. They allow the approximate calculation of solubility by summing up the contributions of relevant atoms or functional groups of compounds using Equation 4:

$$\log S = a_0 + \sum_i a_i n_i \quad (4)$$

The count n_i for atom or functional group type i is obtained from two-dimensional structures of molecules; the contribution a_i for atom or functional group type i is obtained by regression analyses. The n_i can be replaced with solvent accessible surface areas (SAS) to upgrade the models from 2D to 3D.

The methods proposed by Nirmalakhandan *et al.* [43], Suzuki *et al.* [44], Klopman *et al.* [45,46], Hou *et al.* [24], and Wang *et al.* [47] belong to this category. Among these, the Klopman's model and the Hou's model are widely used. The Klopman's model, based on a set of 118 functional groups, leads to a squared correlation coefficient of 0.95 and an average absolute error of 0.50 log unit [46]. In 2004, Hou and co-workers developed an atom contribution model [24]. In this model, 76 atom types were used to classify atoms with different chemical environments, and two correction factors, the hydrophobic carbon and the square of molecular weight, were used to account for the inter-/intra-molecular hydrophobic interactions and bulkiness effect. The contribution coefficients of atom types and correction factors were generated by a multiple linear regression using a learning set consisting of 1290 organic compounds. The obtained linear regression model possesses good statistical significance with r^2 = 0.92, SD = 0.61, and AAE = 0.48. For the 21 tested compounds, a predictive r^2 = 0.88, SD = 0.84, and AAE = 0.52 were achieved. When coming to aqueous solubility prediction of a 21-molecule test set used by Huuskonen, this model achieved a very good accuracy, and it is comparable to or better than most of the published models based on molecular descriptors. In another test, the Hou's model gave better performance than the Klopman's group contribution model in predicting a test set of 120 molecules. Atom/group contribution methods may be the most practical ways of estimating aqueous solubility because they do not need any molecular descriptors based on other theoretical models. Moreover, this class of methods only needs to count the occurrence of functional groups in a molecule, so they are extremely time-saving.

2.4 Prediction accuracies of solubility

Since there are many solubility prediction models available, it is interesting to compare the performance of these models. The comparison of different prediction models is usually assessed by the 21-compound test set used by Huuskonen [13]. It is obvious that this kind of comparison, based on a small test set, is not reliable.

Recently, Kühne and co-workers reported a very interesting study on model comparison and selection [48]. The authors compared the performance of seven models using an in-house data set of 1876 compounds. The data set consists of thoroughly validated experimental values for the water solubility at 25°C of 1876 pure organic chemicals taken from an in-house database. The seven models include the Meylan's model [49], the Klopman's model [46], the Marrero's model [50], the Hou's model [24], the Huuskonen's model [51], the Tetko's model [37], and the Abraham's model [32]. All seven models were developed based on 2-D descriptors, and do not require the melting-point in case of solids.

According to the calculation results, among all these models, only the Hou's model was formally applicable to all compounds, while there were 93 compounds having missing fragments for the Marrero's model. The overall best statistics were achieved by the Meylan's model in term of the predictive squared correlation coefficient (r^2 = 0.83). The prediction performance of the Hou's model (r^2 = 0.82) was pretty close to the Meylan's model. The Abraham's model (r^2 = 0.34) was inferior to all other models, probably because its fragment scheme to predict the linear solvation–energy relationship (LSER) parameters from molecular structure was not ideal. Both the Huuskonen's and the Abraham's models yielded individual prediction errors > 1.0 log units. However, the other five methods also produced individual prediction errors > 3 log units. So according to Kühne's report, we can give a rough estimation of the prediction accuracies for solubility: the best prediction model (the Meylan's model or the Hou's model) can give a standard error < 0.9 log unit.

3. Prediction of human intestinal absorption

Almost all the biological procedures involved in crossing biological membranes have a similar mechanism, such as the drug permeability through the barrier of the gastrointestinal tract, the drug permeability through the blood–brain barrier, and the drug permeability through the Caco-2 monolayers. The major route for the drug permeability through the barrier, passive diffusion, is driven by a concentration gradient. Two types of passive diffusion mechanisms exist: paracellular transport and transcellular transport. In addition to passive diffusion, some molecules can be transported by the active transporters, which include both active carrier systems such as the monocarboxylic acid carrier (which transports salicylic acid) and efflux systems such as *P*-glycoprotein. For both intestinal epithelium and the blood–brain barrier, the

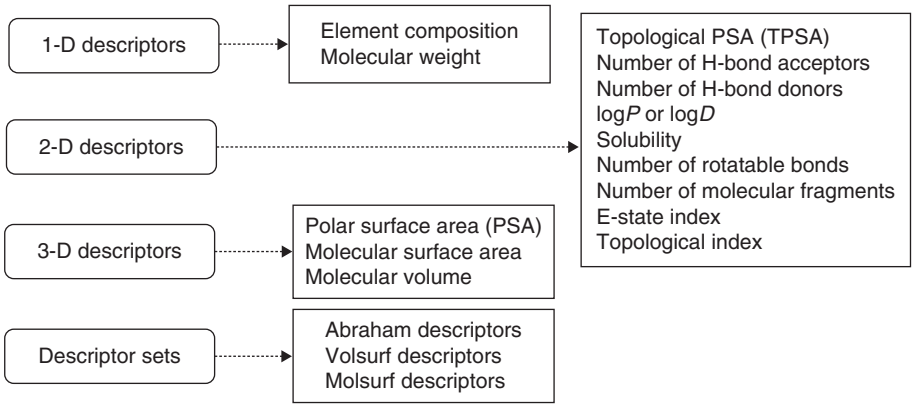

Figure 2. The molecular descriptor used in the predictions of drug transport through biological barriers. Molecular descriptors can be roughly divided into three categories: 1-D (one-dimensional), 2-D, and 3-D descriptors. 1-D descriptors are dependent only on the formula of a molecule; 2-D descriptors are obtained from the connectivity or graph of a molecule; 3-D descriptors contain the 3-D geometric information of a molecule. Descriptor set usually includes a group of 2-D and 3-D molecular descriptors.

transcellular passive diffusion is more important, and thus the prediction of drug absorption and permeability concentrates on this pathway. Here we review the most important transport process: permeability and absorption through the barrier of the gastrointestinal tract. Given the high similarity of all these biological barriers, the methods and the descriptors used for predicting all biological barriers are also similar.

In experiment, HIA is measured by fraction absorption, %FA, which is defined as the total mass absorbed divided by the given dose of the drug. The theoretical prediction of HIA was pioneered by the 'rule-of-five' proposed by Lipinski and co-workers [8]. The rule-of-five defined several criteria for identifying compounds with possible poor absorption and permeability: molecular weight > 500; calculated log*P* > 5 (CLOGP) or > 4.15 (MLOGP); number of hydrogen bond donors (OH and NH groups) > 5; and number of hydrogen bond acceptors (N and O atoms) > 10. Poor absorption and permeation are more likely to occur when any two of the above rules are satisfied. The disadvantage of the rule-of-five is that it can give only a rough classification of molecules, allowing the elimination of only a very limited set.

3.1 *In silico* prediction models for HIA

When predicting HIA, 2-D and 3-D molecular descriptors have generally been used as variables to generate the prediction models [16]. These descriptors define a variety of molecular properties, including lipophilicity, hydrogen bonding ability, molecular bulkiness, etc. Among these molecular descriptors (**Figure 2**), polar surface area (PSA) and apparent partition coefficient (log*D*) may be more important than the others [16,17].

In 1992, van de Waterbeemd and Kansy correlated PSA of a series of CNS drugs to blood–brain partitioning [52]. Since then, PSA has become the most popular descriptor for the prediction of molecular transport properties. In 1997, Palm and co-workers found that an excellent sigmoidal relationship could be established between FA and PSA (r^2 = 0.94) for a set of 20 drugs covering a wide range of %FA values in humans, and concluded that drugs that were completely absorbed (FA > 90%) had a PSA ≤ 61 Å^2, while drugs that were < 10% absorbed had a PSA ≥ 140 Å (**Figure 3A**) [53]. Hou and co-workers checked the relationships between topological polar surface area (TPSA) and %FA for 553 molecules, and a much poorer correlation (r^2 = 0.49) was observed (**Figure 3B**). According to the results reported by Hou and co-workers, applying the value of 61 Å^2, 230 compounds could be identified as possibly being well-absorbed. In these 230 compounds, 47 have an intestinal absorption < 90% and 17 < 80%. For the 266 compounds with a TPSA > 61 and < 140 Å^2, 165 compounds have an intestinal absorption > 90% and five compounds < 10%. It is clear that the performance of the TPSA criterion is not reliable to identify poor absorption or good absorption, and HIA is certainly not only determined by PSA or TPSA. The discrepancy between the two models may be caused by the use of a very limited number of data set in Palm's model, while Hou's model used a much larger and more diverse data set. The similar observation of the poor performance of PSA for predicting drug absorption based on a diverse data set has been reported by others [54]. For example, Grass and Sinko have described the difficulty in using PSA as a sole predictor of human absorption for an extensive data set in the iDEA™ database [54].

The distribution coefficient, log*D*, is also a very important descriptor for the HIA prediction. The hydrophobic parameters (log*P* or log*D*) have long been known to be

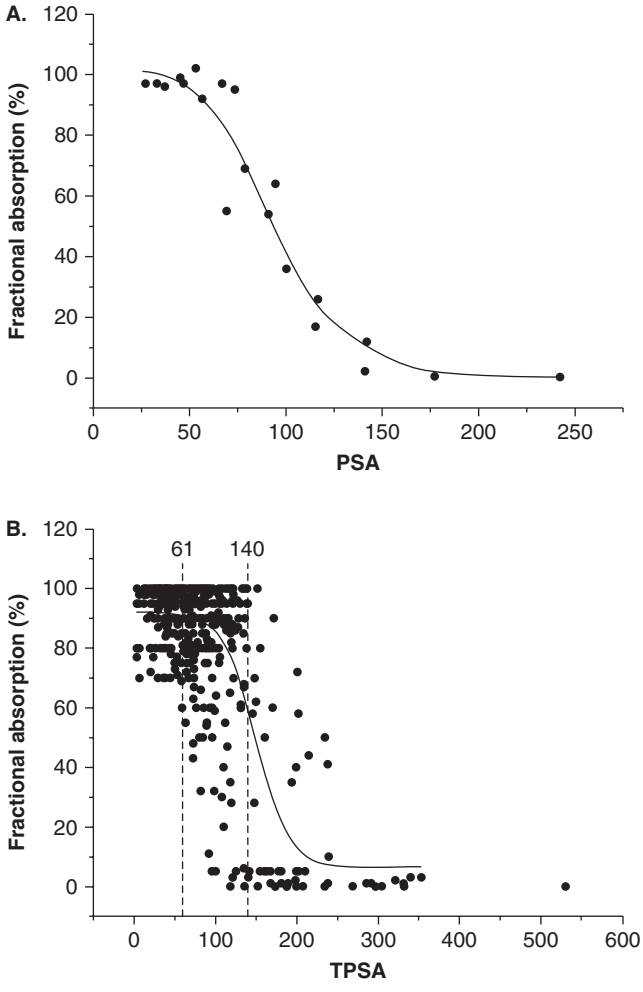

Figure 3. A. The polar surface area for 20 drugs [53]; and B. The topological polar surface area for 553 drugs, versus the experimental human fractional absorption [55].

important for membrane permeation. Hou and co-workers studied the linear correlations between %FA and log*D* at pH = 6.5, and a correlation with r^2 = 0.40 could be observed, which is better than that between %FA and log*P* [55]. In most studies, researchers like to use log*P* instead of log*D* because log*P* is easier to compute. But log*D* is undoubtedly more effective than log*P* in the prediction of membrane permeability. Furthermore, log*D* cannot be replaced by any other descriptors. Recently, Hou and co-workers studied the impact of 10 molecular descriptors for classifying the compounds into good and poor HIA absorption classes [56]. Among these 10 descriptors, TPSA and predicted apparent octanol-water distribution coefficient at pH 6.5 ($\log D_{6.5}$) showed better classification performance than the others.

These two important descriptors, PSA (or TPSA) and log*D*, are usually included in many HIA prediction models. Certainly, other descriptors are necessary to generate more reliable prediction models; for example, the prediction model developed by Hou and co-workers (Equation 5) [55]:

$$\%FA = 97.12 - 11.48N_{\text{rule-of-five}} - 8.99 < 0.05 - \log D_{6.5} > 0.15 < \text{TPSA} - 49.41 > +0.17(\log D_{6.5})^2 + 3.76 < n_{\text{HBD}} - 7 > \quad (n = 435, r^2 = 0.76, SD = 12.70, F = 277.59) \quad (5)$$

In Equation 5, $N_{\text{rule-of-five}}$ is the number of violations of rule-of-five, and N_{HBD} is the H-bond donor count. The spline terms used in Equation 5 are denoted with angled brackets. For example, $< f(x) - a >$ is equal to zero if the value of $f(x) - a$ is negative; otherwise, it is equal to $f(x) - a$. The regression with splines allows the incorporation of features that do not have a linear effect over their entire range. In Equation 5, the threshold value of TPSA is about 50 Å, demonstrating that higher TPSA values produce low permeation, while the effect takes effect only when the PSA is > 50 Å^2. A spline model for log $D_{6.5}$ is also included in the prediction models. A threshold of 0.05 was found for log $D_{6.5}$, which means that log $D_{6.5}$ values < 0.05 produce low permeation. The interpretation of the n_{HBD} term is not very straightforward. This term indicates that n_{HDB} is unfavorable for HIA when it is > 7. This term may be used for the neutralization of the strong effect of TPSA and $N_{\text{rule-of-5}}$.

Besides the traditional descriptors, some other descriptor sets were applied in the HIA predictions, such as the Abraham descriptors [57], the Volsurf descriptors [58], and the Molsurf descriptors [59]. Equation 6 is the prediction model based on the Abraham descriptors to model the HIA data of 169 drugs reported by Zhao and co-workers [57]. The obtained model possesses good correlation and external prediction ability. The stepwise regression analysis showed that the two dominated descriptors are $\sum a_2^H$ and $\sum \beta_2^H$, in good agreement with previous work that suggested hydrogen-bond donors and acceptors, or polar molecular surface, were good descriptors to model HIA.

$$\%FA = 90 + 2.11R_2 + 1.70\pi_2^H - 20.7\sum a_2^H - 22.3\sum \beta_2^H + 15.0V_x \quad (n = 38, r^2 = 0.83, q^2 = 0.75, SD = 16\%, F = 31) \quad (6)$$

3.2 Prediction accuracies of HIA

The reliable evaluation of the prediction accuracy of a model should be based on a precise and extensive data set. The data sets used by many of the previous models for the predictions of HIA include only a small number of compounds (20 – 40) [17]. Based on the limited data set, the prediction accuracy of the prediction model cannot be reliably guaranteed. In 2007, Hou and co-workers reported

a fairly large data set for HIA, which includes 647 drug and drug-like molecules collected from a variety of literature sources [55]. Among these 647 molecules, 578 are believed to be transported by passive diffusion. Based on the data set, Hou developed a set of prediction models for HIA [55].

The theoretical correlation model for a training set of 455 compounds was proposed by using the genetic function approximation (GFA) technique. The model was able to predict the fractional absorption with an $r^2 = 0.71$ and an average absolute error of 11.2% for the training set. Moreover, it achieved an $r^2 = 0.81$ and an average absolute error of 7.3% for a 98-compound test set. So according to Hou's report, we can give a rough estimation of the prediction accuracies for HIA: the best prediction model can give an average absolute error less than 10% (7.3% for the tested compounds).

Based on the same data set, Hou and co-workers reported a classification model based on the recursive partitioning (RP) technique to classify the compounds into poor (%FA ≤ 30%, defined as class I) or good (%FA > 30%, defined as class II) HIA [55]. The obtained model had very good classification performance on the training set, and it could correctly identify 95.9% (71/74) of the compounds in class I and 96.1% (391/407) of the compounds in class II. It was encouraging that the performance on the test set was also very satisfactory. The test set included five compounds in class I and 93 compounds in class II. All five compounds in class I were correctly classified, and only three compounds in class II were not correctly identified.

As a comparison, Hou and co-workers studied the performance of a support vector machine (SVM) to classify compounds into high or low fractional absorption [56]. The best SVM classifier could give satisfactory predictions for the training set (97.8% for class I and 94.5% for class II). Moreover, 100% of the compounds in class I and 97.8% of the compounds in class II in the external test set could be correctly classified. The total number of misclassified number was decreased from 22 of RP to 15 of SVM. It seems that the SVM classifier gave more reliable predictions than the RP model, based on either the prediction for the training set or that for the test set. It is obvious that the classification model based on SVM has very good capability to discriminate the well-absorbed compounds and the poorly absorbed compounds.

4. Prediction of oral bioavailability

Oral bioavailability (F) is defined as the fraction of the ingested dose of a drug that is available to the systematic circulation following oral administration. The oral bioavailability of a drug is usually < 100%, considering degradation or metabolism of the drug prior to absorption, incomplete absorption and first-pass metabolism. Compared with the prediction of HIA, the prediction of oral bioavailability is considerably more challenging because bioavailability is a complex function of many biological and physicochemical factors, such as dissolution in the gastrointestinal tract, intestinal membrane permeation, intestinal and hepatic first-pass metabolism, and even the dosage form. Furthermore, these factors may vary from patient to patient, and even vary in the same patient over time. Whether a drug is taken with or without food will affect absorption, and other drugs taken concurrently may alter absorption and first-pass metabolism. Moreover, disease states affecting liver metabolism or gastrointestinal function will also have an effect.

4.1 *In silico* prediction models for oral bioavailability

In the last several years, several prediction models of oral bioavailability based on QSPR analysis have been reported [60-63].

In 2000, Andrews and co-workers developed a regression model to predict oral bioavailability [60]. Compared to the Lipinski's rule-of-five, the false negative predictions were reduced from 5% to 3%, while the false positive predictions decreased from 78% to 53%. The model could achieve a relatively good correlation ($r^2 = 0.71$) for the training set. But when 80/20 cross-validation was applied, the correlation was decreased to $q^2 = 0.58$.

Recently, Wang and co-workers reported another regression model to predict oral bioavailability using the counts of functional groups as descriptors [61]. A genetic algorithm was employed to find the prediction models with the best combination of functional groups. The final models include 42 functional groups and two other molecular descriptors: molecular refractivity and rule-of-five. The mean r^2 and mean *rmse* for the 20 best models were 0.55 and 21.9%, respectively. For the 90/10 cross-validation, the mean r^2 and mean *rmse* for the test sets were 0.42 and 24.6%, respectively. Similar to the model reported by Andrews [60], the q^2 of this model was 0.13 lower than the r^2, although the *rmse* was only marginally increased.

The classification models for predicting oral bioavailability have also been proposed; for example, the model developed by Yoshida and co-workers [62]. The Yoshida's model was developed based on a set of physiochemical parameters, including distribution coefficient at pH = 6.5 ($\log D_{6.5}$), $\Delta \log D$ ($\log D_{6.5} - \log D_{7.4}$), and 15 functional groups related to well-known metabolic pathways [62]. The ORMUCS (ordered multicategorical classification method using the simplex technique) method was applied to assign the oral bioavailability into one of four classes. In the leave-one-out cross-validation tests, an average of 67% of the drugs were correctly classified. The predictive power of the model was evaluated using a separate test set of 40 compounds, of which 60% (95% within the same class) were correctly classified.

4.2 Prediction accuracies of oral bioavailability

It is obvious that the available prediction models cannot give reliable estimations for oral bioavailability. According

to the publications reported by Andrews *et al.* and Wang *et al.* [60,61], we can give a rough estimation of prediction accuracies for oral bioavailability: the prediction models for oral bioavailability can give a *rmse* for prediction > 20% (20.40% for the Andrew's model and 24.6% for the Wang's model).

Now, all prediction models for oral bioavailability are developed based on molecular descriptors. It is possible that the hepatic metabolism cannot be effectively explained by these molecular properties, and thus highly metabolized compounds may not be well predicted by these simple descriptor-based models. Recently, Hou *et al.* reported an analysis of a database of human oral bioavailability for 768 chemical compounds [64]. The correlations between several important molecular descriptors and human oral bioavailability were investigated and compared with the earlier work reported by Veber *et al.* [65]. The analysis showed that the percentages of compounds meeting the criteria based on molecular descriptors did not distinguish compounds with poor oral bioavailability from those with acceptable values, which may suggest that no simple model based on molecular descriptors can be used as a general filter to predict oral bioavailability with high confidence. The performance of these rules based on molecular descriptors in the prediction of HIA is obviously much better than that of oral bioavailability in term of false positive rate. Therefore, the prediction models based on molecular descriptors can give good predictions for human intestinal absorption, but cannot give reliable predictions for oral bioavailability.

5. Expert opinion

The progress in computational modeling of solubility, HIA and oral bioavailability is briefly reviewed here. Significant effort continues in modeling these three important ADME properties, but much work is still necessary to make predictions more reliable and accurate to significantly impact upon the drug discovery process. For the three ADME properties discussed here, only passive human intestinal absorption can be predicted with relatively good accuracy. Moreover, it should be pointed out that the prediction accuracy of the new models does not appear to have demonstrated much progress, although many prediction models have been developed in the past few years. For example, for the seven solubility models compared by Kühne and co-workers [48], the best model developed by Meylan *et al.* [49] is a relatively old one. For the regression models for predicting oral bioavailability, the new model developed by Wang *et al.* in 2007 [47] does not show better performance than the Andrews's model developed in 2000, although a much larger data set was used in Wang's model [60]. So the prediction accuracy of the *in silico* models is still the biggest challenge we are facing now.

Undoubtedly, the lack of extensive and reliable experimental data are an important reason to hinder the development of reliable ADME prediction models. It is particularly true for the *in vivo* oral bioavailability and human intestinal absorption data, which are usually collected for drugs or drug candidates in clinic trials. In addition, these data may show significant variability from one source to another [55,64]. The largest pharmaceutical companies have developed large in-house databases containing consistently measured compound properties. For example, Veber *et al.* reported an analysis of bioavailability in rat on a data set of > 1100 compounds studied at GlaxoSmithKline [65]; Andrew *et al.* published a study of oral bioavailability for 591 structures from Glaxo Wellcome's internal database [60]. However, these data are usually not available publicly for the scientific community. Therefore, many models are still developed based on small historical data sets taken from the literature. Encouragingly, several developments have been achieved with the availability of large data sets in the recent years. For example, in 2007, two extensive data sets for HIA and oral bioavailability were reported by Hou and co-workers [55,64], which give more opportunities for the development of reliable prediction models in the future. Certainly, further developments on the availability of ADME data for the public domain are still necessary.

Many models are reported to predict the aqueous solubility of drug-like compounds; but most of the available solubility prediction models may only work reliably on non-charged compounds, because in the training process of these models the effect of ionization was not explicitly considered. It is well known that solubility is strongly dependent on many factors, such as the effect of ionization and the crystal size of the solute. Unfortunately, only limited publications report the development of solubility prediction models by considering these important aspects of solubility. For example, in 2006, Hansen and co-workers employed the Henderson-Hasselbalch (HH) equation for the prediction of pH-dependent aqueous solubility of drugs and drug candidates [66]. The intrinsic solubility was developed based on artificial neural networks and the pKa was predicted by the Marvin software developed by ChemAxon. For a data set of 27 drugs, the experimentally determined pH–solubility curves were used for the validation of the combined pH-dependent model, and that model could give a mean *rmse* of 0.79 log*S* units. We expect that in the coming years more sophisticated models will emerge that more tightly integrate measures of ionization, crystal packing, and salt effects.

For oral bioavailability, no model can yet give reliable predictions. The poor prediction of oral bioavailability is caused primarily by poor prediction for the first-pass metabolism. Predicting oral bioavailability now follows similar strategies for predicting most of the other ADME properties: by generating molecular descriptors of molecular structures in the data set and developing the prediction models. The strategy works well for the prediction of many ADME properties, but obviously does not work well for the prediction of oral bioavailability because the molecular descriptors in

current use cannot effectively characterize the first-pass metabolism. Therefore, introducing new rules or descriptors to model the first-pass metabolism is the most important requirement to be met in order to develop accurate prediction models for oral bioavailability. With the continually increased data of oral bioavailability, we may find better 'substructure-specific rules' directly related to different metabolism pathways in the future. Using these substructure-specific rules as descriptors, the predictions of oral bioavailability may be improved. Another research direction for the prediction of oral bioavailability is to develop separated prediction models for different components involved in oral bioavailability, including passive transcellular transport, paracellular transport, carrier-mediated transport, and first-pass metabolism, and then integrate them together. At present, the development of an integrated model is extremely difficult, sometimes impossible, because the predictions for some mechanisms involved in oral bioavailability are unreliable.

Another possible way to improve the prediction accuracy of the ADME properties is to develop consensus models by combining two or more models for the same property together. Actually, the concept of 'consensus model' has long been introduced into the ADME predictions. For example, researchers at Bio-Rad Laboratories, Inc. have introduced the consensus score to the KnowItAll ADME/Tox software system [67], and found that the employment of multiple complementary models for the same ADME-Tox end point in consensus modeling provides a greater accuracy than that of any single model. Recently, Abshear *et al.* validated the performance of four models in KnowItAll for predicting the intrinsic solubility of 113 diverse organic compounds [68]. For predicted aqueous solubility of 113 compounds, four individual models gave absolute average errors of 0.314, 0.422, 0.327 and 0.324 log units, respectively. By combining these four individual models, the consensus model gave a absolute average error of 0.257 log units.

Finally, it is worth emphasizing that although a lot of commercial software systems are available for predicting the ADME properties, we still know little about the performance of these software systems on specific ADME property. It is well recognized among the community that despite the best efforts of their developers, some models that are assumed to be extensively validated do not have the desired performance [69]. More comparisons of the prediction performance of different software systems are urgently required in order to guide scientists to choose the most appropriate model in the drug discovery process.

Declaration of interest

The authors state no conflict of interest and have received no payment in preparation of this manuscript.

Bibliography

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

1. Hou TJ, Xu XJ. Recent development and application of virtual screening in drug discovery: An overview. Curr Pharm Des 2004;10:1011-33
2. Segall MD, Beresford AP, Gola JMR, et al. Focus on success: using a probabilistic approach to achieve an optimal balance of compound properties in drug discovery. Expert Opin Drug Metab 2006;2:325-37
3. Beresford AP, Selick HE, Tarbit MH. The emerging importance of predictive ADME simulation in drug discovery. Drug Discov Today 2002;7:109-16
4. Kennedy T. Managing the drug discovery/development interface. Drug Discov Today 1997;2:436-44
5. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 2004;3:711-5

• **The analysis of attrition rate of drugs, based on relatively new data.**

6. Li AP. Screening for human ADME/Tox drug properties in drug discovery. Drug Discov Today 2001;6:357-66
7. Bevan CD, Lloyd RS. A high-throughput screening method for the determination of aqueous drug solubility using laser nephelometry in microtiter plates. Anal Chem 2000;72:1781-7
8. Lipinski CA, Lombardo F, Dominy BW, et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 1997;23:3-25

•• **One of the most important review of solubility and permeability predictions and their importance to the pharmaceutical industry.**

9. Wienkers LC, Heath TG. Predicting in vivo drug interactions from in vitro drug discovery data. Nat Rev Drug Discov 2005;4:825-33
10. Crespi CL, Miller VP, Penman BW. Microtiter plate assays for inhibition of human, drug-metabolizing cytochromes P450. Anal Biochem 1997;248:188-90
11. Cohen LH, Remley MJ, Raunig D, et al. In vitro drug interactions of cytochrome P450: An evaluation of fluorogenic to conventional substrates. Drug Metab Dispos 2003;31:1005-15
12. van de Waterbeemd H, Gifford E. ADMET in silico modelling: Towards prediction paradise? Nat Rev Drug Discov 2003;2:192-204

•• **An excellent review of *in silico* ADME predictions.**

13. Delaney JS. Predicting aqueous solubility from structure. Drug Discov Today 2005;10:289-95

• **An assessment of the theoretical models for solubility predictions.**

14. Johnson SR, Zheng WF. Recent progress in the computational prediction of aqueous solubility and absorption. AAPS J 2006;8:E27-40
15. Norinder U, Bergstrom CAS. Prediction of ADMET properties. ChemMedChem 2006;1:920-37
16. Hou TJ, Wang JM, Zhang W, et al. Recent advances in computational prediction of drug absorption and permeability in drug discovery. Curr Med Chem 2006;13:2653-67

• **A valuable review of *in silico* predictions of drug absorption.**

17. Hou TJ, Li YY, Zhang W, et al. Recent developments of in silico predictions of intestinal absorption and oral

bioavailability. Comb Chem High Throughput Screen 2008; In press

18. Fagerholm U. Prediction of human pharmacokinetics: gastrointestinal absorption. J Pharm Pharmacol 2007;59:905-16
19. Norinder U, Haeberlein M. Computational approaches to the prediction of the blood-brain distribution. Adv Drug Deliv Rev 2002;54:291-313
20. Ekins S, Ecker GF, Chiba P, et al. Future directions for drug transporter modelling. Xenobiotica 2007;37:1152-70
21. Jolivette LJ, Ekins S. Methods for predicting human drug metabolism. Adv Clin Chem 2007;43:131-76
22. Gola J, Obrezanova O, Champness E, et al. ADMET property prediction: The state of the art and current challenges. Qsar Comb Sci 2006;25:1172-80
23. 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;94:2327-43
24. Hou TJ, Xia K, Zhang W, et al. ADME evaluation in drug discovery. 4. Prediction of aqueous solubility based on atom contribution approach. J Chem Inf Comp Sci 2004;44:266-75
25. Jorgensen WL, Duffy EM. Prediction of drug solubility from structure. Adv Drug Deliv Rev 2002;54:355-66
26. Available from: www.fda.gov/cder/OPS/BCS_guidance.htm
27. Fagerholm U. Evaluation and suggested improvements of the Biopharmaceutics Classification System (BCS). J Pharm Pharmacol 2007;59:751-7
28. Yazdanian M, Briggs K, Jankovsky C, et al. The 'high solubility' definition of the current FDA Guidance on Biopharmaceutical Classification System may be too strict for acidic drugs. Pharm Res 2004;21:293-9
29. Yalkowsky SH, Valvani SC. Solubility and partitioning.1. Solubility of non-electrolytes in water. J Pharm Sci 1980;69:912-22
30. Jain N, Yalkowsky SH. Estimation of the aqueous solubility I: Application to organic nonelectrolytes. J Pharm Sci 2001;90:234-52
31. Huuskonen J, Salo M, Taskinen J. Aqueous solubility prediction of drugs based on molecular topology and neural network modeling. J Chem Inf Comp Sci 1998;38:450-6
32. Abraham MH, Le J. The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship. J Pharm Sci 1999;88:868-80
33. Jorgensen WL, Duffy EM. Prediction of drug solubility from Monte Carlo simulations. Bioorg Med Chem Lett 2000;10:1155-8
34. McElroy NR, Jurs PC. Prediction of aqueous solubility of heteroatom-containing organic compounds from molecular structure. J Chem Inf Comp Sci 2001;41:1237-47
35. McFarland JW, Avdeef A, Berger CM, et al. Estimating the water solubilities of crystalline compounds from their chemical structures alone. J Chem Inf Comp Sci 2001;41:1355-9
36. Liu RF, Sun HM, So SS. Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. 2. Blood-brain barrier penetration. J Chem Inf Comp Sci 2001;41:1623-32
37. Tetko IV, Tanchuk VY, Kasheva TN, et al. Estimation of aqueous solubility of chemical compounds using E-state indices. J Chem Inf Comp Sci 2001;41:1488-93
38. Engkvist O, Wrede P. High-throughput, in silico prediction of aqueous solubility based on one- and two-dimensional descriptors. J Chem Inf Comp Sci 2002;42:1247-9
39. Yan AX, Gasteiger J. Prediction of aqueous solubility of organic compounds based on a 3D structure representation. J Chem Inf Comp Sci 2003;43:429-34
40. Butina D, Gola JMR. Modeling aqueous solubility. J Chem Inf Comp Sci 2003;43:837-41
41. Goller AH, Hennemann M, Keldenich J, et al. In silico prediction of buffer solubility based on quantum-mechanical and HQSAR- and topology-based descriptors. J Chem Inf Model 2006;46:648-58
42. Huuskonen J. Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology. J Chem Inf Comp Sci 2000;40:773-7

• **An important work on prediction of aqueous solubility.**

43. Nirmalakhandan NN, Speece RE. Prediction of aqueous solubility of organic chemicals based on molecular structure. 2. Application to Pnas, Pcbs, Pcdds, etc. Environ Sci Technol 1989;23:708-13
44. Suzuki T. Development of an automatic estimation system for both the partition-coefficient and aqueous solubility. J Comput Aid Mol Des 1991;5:149-66
45. Klopman G, Wang S, Balthasar DM. Estimation of aqueous solubility of organic molecules by the group contribution approach: application to the study of biodegradation. J Chem Inf Comp Sci 1992;32:474-82
46. Klopman G, Zhu H. Estimation of the aqueous solubility of organic molecules by the group contribution approach. J Chem Inf Comp Sci 2001;41:439-45
47. Wang JM, Krudy G, Hou TJ, et al. Development of reliable aqueous solubility models and their application in druglike analysis. J Chem Inf Model 2007;47:1395-404
48. Kühne R, Ebert RU, Schuurmann G. Model selection based on structural similarity - Method description and application to water solubility prediction. J Chem Inf Model 2006;46:636-41

•• **Comparison of several models for predicting aqueous solubility.**

49. Meylan WM, Howard PH. Estimating log P with atom/fragments and water solubility with log P. Perspect Drug Discov 2000;19:67-84
50. Marrero J, Gani R. Group-contribution-based estimation of octanol/water partition coefficient and aqueous solubility. Ind Eng Chem Res 2002;41:6623-33
51. Huuskonen J. Estimation of water solubility from atom-type electrotopological state indices. Environ Toxicol Chem 2001;20:491-7
52. van de Waterbeemd H, Kansy M. Hydrogen-bonding capacity and brain penetration. Chimia 1992;46:299-303
53. Palm K, Stenberg P, Luthman K, et al. Polar molecular surface properties predict the intestinal absorption of drugs in humans. Pharm Res 1997;14:568-71
54. Grass GM, Sinko PJ. Effect of diverse datasets on the predictive capability of ADME models in drug discovery. Drug Discov Today 2001;6:S54-61
55. Hou TJ, Wang JM, Zhang W, et al. ADME evaluation in drug discovery. 7. Prediction of oral absorption by correlation and

classification. J Chem Inf Model 2007;47:208-18

•• **Used the largest public community data set for for modeling human oral absorption.**

56. Hou TJ, Wang JM, Li YY. ADME evaluation in drug discovery. 8. The prediction of human intestinal absorption by a support vector machine. J Chem Inf Model 2007;47:2408-15

57. 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;90:749-84

58. Cruciani C, Crivori P, Carrupt PA, et al. Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. J Mol Struc Theochem 2000;503:17-30

59. Norinder U, Osterberg T, Artursson P. Theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics. Eur J Pharm Sci 1999;8:49-56

60. Andrews CW, Bennett L, Yu LX. Predicting human oral bioavailability of a compound: Development of a novel quantitative structure–bioavailability relationship. Pharm Res 2000;17:639-44

61. Wang JM, Krudy G, Xie XQ, et al. Genetic algorithm-optimized QSPR models for bioavailability, protein binding, and urinary excretion. J Chem Inf Model 2006;46:2674-83

62. Yoshida F, Topliss JG. QSAR model for drug human oral bioavailability. J Med Chem 2000;43:2575-85

63. Turner JV, Glass BD, Agatonovic-Kustrin S. Prediction of drug bioavailability based on molecular structure. Anal Chim Acta 2003;485:89-102

64. Hou TJ, Wang JM, 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;47:460-3

• **An extensive database of oral bioavailability for the scientific community.**

65. Veber DF, Johnson SR, Cheng HY, et al. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 2002;45:2615-23

66. Hansen NT, Kouskoumvekaki I, Jorgensen FS, et al. Prediction of pH-dependent aqueous solubility of druglike molecules. J Chem Inf Model 2006;46:2601-9

67. Available from: www.bio-rad.com

68. Abshear T, Banik GM, D'Souza ML, et al. A model validation and consensus building environment. Sar Qsar Environ Res 2006;17:311-21

• **An interesting study showing the advantage of the consensus model.**

69. Stouch TR, Kenyon JR, Johnson SR, et al. In silico ADME/Tox: why models fail. J Comput Aid Mol Des 2003;17:83-92

Affiliation

Tingjun Hou[†1] & Junmei Wang[2]
[†]Author for correspondence
[1]University of California at San Diego, Department of Chemistry and Biochemistry, Center for Theoretical Biological Physics, 9500 Gilman Drive, La Jolla, CA 92093-0359, USA
Tel: +1 858 822 4596; Fax: +1 858 822 4236; E-mail: tingjunhou@hotmail.com
[2]The University of Texas Southwestern Medical Center, Department of Pharmacology, 5323 Harry Hines Blvd Dallas, TX 75390, USA