# *Current methodology for the assessment of ADME-Tox properties on drug candidate molecules*

REVIEW

✍ Ania de la Nuez, Rolando Rodríguez

División de Química Física, Centro de Ingeniería Genética y Biotecnología, CIGB
Ave. 31 e/ 190 y 158, AP 6162, CP 10 600, Playa, Ciudad de La Habana, Cuba
E-mail: ania.delanuez@cigb.edu.cu

***ABSTRACT***

Absorption, distribution, metabolism, excretion and toxicity (ADME-Tox) properties should be considered to develop a new drug, because they are the main cause of failures for candidate molecules in drug design. The early evaluation of these properties durin drug design could save time and money. Physicochemical properties, tridimensional (3D) structural information, and mathematical methods can be combined to develop an *in silico* approach to predict ADME-Tox properties. Some of these *in silico* methods can be used to predict these properties in a large number of compounds at an early stage in drug design, but there is no general methodology for the computer prediction of ADME-Tox properties. In this review we summarize some of the models and available programs to predict ADME-Tox properties.

*Keywords*: ADME, absorption, distribution, metabolism, excretion, toxicity, drug design, QSAR, molecular descriptors, pharmacokinetic

*Biotecnología Aplicada 2008;25:97-110*

***RESUMEN***

**Metodología actual para la evaluación de propiedades ADME-Tox de moléculas candidatas a medicamentos.** Las propiedades ADME-Tox (absorción, distribución, metabolismo, excreción y toxicidad) deben tenerse en cuenta en el desarrollo de nuevos medicamentos. Estas propiedades son la principal causa del fallo de moléculas candidatas durante el diseño de medicamentos. La evaluación de estas propiedades en etapas tempranas del diseño de medicamentos permite ahorrar tiempo y gastos. Propiedades físico-químicas, información de estructuras tridimensionales y métodos matemáticos pueden ser combinados para desarrollar modelos *in silico* para predecir propiedades ADME-Tox. Algunos de estos métodos *in silico* se pueden usar para predecir estas propiedades a un gran número de compuestos en etapas tempranas del diseño de medicamentos, pero no hay una metodología general para la predicción de las propiedades ADME-Tox. Esta revisión resume algunos modelos y programas disponibles para predecir propiedades ADME-Tox.

*Palabras clave*: ADME, absorción, distribución, metabolismo, excreción, toxicidad, diseño de fármacos, QSAR, descriptores moleculares, farmacocinética

## *I*ntroduction

The fact that a chemical compound could be active and selective does not necessarily make it an attractive candidate for drug development. There are certain properties that make a drug different from other compounds. An appropriate concentration of the drug must circulate in the body for a reasonable length of time to achieve a desired beneficial effect with a minimum of adverse effects. For this process, oral drugs have to dissolve or suspend in the gastrointestinal tract and be absorbed through the gut wall, then they pass the liver to reach the blood stream. From there, the drug will be distributed to various tissues and organs and finally binds to its molecular target and exert its desired action. The drug is then subjected to hepatic metabolism followed by its elimination as bile or via the kidneys.

Several pharmacokinetics properties are involved in this mechanism. Bioavailability depends on absorption and liver first-pass metabolism. The volume of its distribution, together with its clearance rate, determines the half-life of a drug and therefore its dosage. Poor biopharmaceutical properties, such as poor aqueous solubility and slow dissolution rate can lead to poor oral absorption and hence low oral bioavailability.

Promising drugs candidates often fail because of unsatisfactory absorption, distribution, metabolism, excretion, and toxicity (ADME-Tox) properties as reported for about 63% of the compounds proposed for pre-clinical development [1].

The conversion of active compounds into qualified clinical candidates has proved to be a challenge. At the molecular level, a coordinated system of transporters, channels, receptors and enzymes act as gatekeepers to foreign molecules affecting the ADME-Tox properties of a given molecule in very different ways. A fast evaluation of ADME properties at the early stages of drug discovery could save both time and money.

This review summarizes some *in silico* approaches to predict ADME-Tox properties.

## *A*DME-Tox at the early stage

To identify compounds that have good pharmacokinetic and toxicological profiles, ADME-Tox studies should be started as early as possible in the discovery process [2-4] enabling all properties to be optimized simultaneously.

Traditional ADME-Tox assays were designed as detailed experimental approaches to characterize a process by its underlying mechanisms, *e.g.* the characterization of the metabolism or transport of a compound would involve investigations at multiple concentrations and time points [5, 6] with a throughput of a few compounds per week. However, these appro-

1. Lipper RA. How can we optimize selection of drug development candidates from many compounds at the discovery stage?. Mod Drug Discovery 1999;2:55.

2. Thompson TN. Optimization of metabolic stability as a goal of modern drug design. Med Res Rev 2001;21:412-49.

3. Smith DA, van de Waterbeemd H. Pharmacokinetics and metabolism in early drug discovery. Curr Opin Chem Biol 1999;3:373-8.

4. Kerns EH. High throughput physicochemical proûling for drug discovery. J Pharm Sci 2001;90:1838.

5. Adachi Y, Suzuki H, Sugiyama Y. Comparative studies on *in vitro* methods for evaluating *in vivo* function of MDR1 P-glycoprotein. Pharm Res 2001;18:1660-8.

6. Andrews J, Abd-Ellah MF, Randolph NL, Kenworthy KE, Carlile DJ, Friedberg T, Houston JB. Comparative study of the metabolism of drug substrates by human cytochrome P450 3A4 expressed in bacterial, yeast and human lymphoblastoid cells. Xenobiotica 2002;32:937-47.

✍ Corresponding author

aches might still be useful as secondary assays during optimization or preclinical development.

New formats of *in vitro* ADME-Tox assays serving early discovery have been developed from traditional assays through protocol simplification and making use of technological advancements. Examples include the measurement of metabolic stability or permeability of a compound at a single concentration and single time point [7] or the evaluation of the cytochrome P450 (CYP450) inhibition potential of a compound at a single concentration, instead of determining the 50% inhibitory concentration (IC50) or inhibition constant value (Ki) [8], high-throughput solubility assays (Turbidometry [9], laser nephelometry[10]), permeability assay (Caco-2 cells [11, 12], Madin-Darby canine kidney (MDCK) cells [13], parallel artificial membrane permeation assay (PAMPA) [14, 15]), and metabolism assays (hepatocytes [16], S9 fractions [17], recombinant enzymes [18]). On the other hand, computer models have appeared to further expand the ADME-Tox tools. These models are promising as early screening tools for drug candidates and for designing more successful combinatorial or *in silico* libraries, where there is a very small amount of the compound or no compound at all and a computational approach is then the only option for collecting this information.

Commercially available tools for calculating physicochemical properties and ADME-Tox-related parameters are often used [19-22]. However, the predictions are far from perfect, and the results obtained from such tools have to be interpreted with great care [23]. The use of generic models can only be recommended if they have been validated for a particular project. If new compound classes outside of the training sets are evaluated, the result can be very misleading.

The optimal approach for the ADME-Tox support of discovery will be one that uses both *in vitro* and *in silico* experiments in a complementary way ensuring that ADME-Tox is used and considered at almost every stage of the discovery process, from hit identification to lead optimization (Figure 1) [19].

In the hit identification stage, the primary goal of the *in silico* ADME-Tox models is to identify compounds or series of compounds with at least acceptable drug-like properties that are then disregarded. Another goal is to identify potential weaknesses and liabilities in the selected series highlighting the issues that will be focused in the improvement/optimization efforts. In the lead identification stage, the objective is to identify a small number of chemical series with the activity, selectivity and drug-like properties required for a potential candidate.

The application of *in silico* ADME-Tox should focus on predictions of chemical modifications of compounds that will improve ADME-Tox properties. *In vitro* assays are used to measure the ADME-Tox properties of the newly synthesized compounds. This information is valuable for the refinement of the *in silico* ADME-Tox models.

Similar to lead identification, the lead optimization on ADME-Tox properties consist of an iterative workflow, starting from *in silico* prediction, to chemical synthesis, to experimental testing and confirmation, and to model refinement.

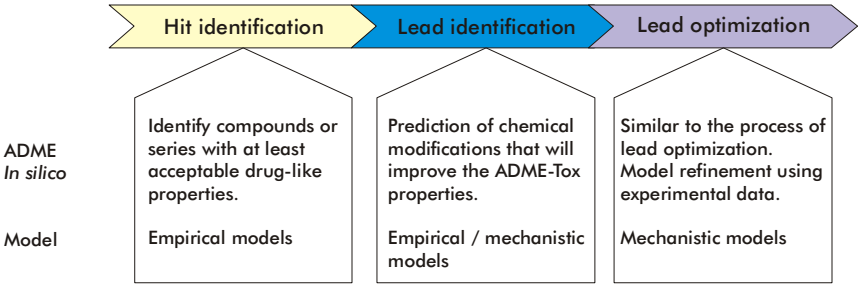


Figure 1. *In silico* ADME-Tox predictions at different stage of drug design. The purpose of these predictions is different in each stage. The models used at the beginning of the drug design have to work fast to evaluate a large number of compounds, and at the later stages, they have to be more accurate.

## *In silico* ADME-Tox models in drug discovery

The current available computer tools to predict ADME properties can be classified in data-based and structure-based approaches.

Data-based approaches include quantitative structure-activity relationship (QSAR) [24] and quantitative structure-property relationship (QSPR) [25]. These approaches use statistical tools to explore the linear or nonlinear relationship between certain structural descriptors and observed parameters of a particular ADME-Tox property [26-29]. Various kinds of quantitative descriptors based on 2-dimensional or 3-dimensional molecular structures have been proposed, including fragment, topological and global physicochemical descriptors. On the other hand, to relate the target property to the descriptors, linear methods (*e.g.*, multiple linear regression and partial least squares) and non-linear methods (e.g., feed-forward artificial neural network) have been applied for multivariate analysis [30].

Data-mining and machine-learning methods originally developed and used in other fields are now also successfully being used for this purpose. Examples of such methods include neural networks [26, 31], self-organizing maps [32], recursive partitioning [33, 34] and support vector machines [35, 36].

There are also data-based methods that do not use an explicit mathematical model. They are mostly based on molecular similarity/dissimilarity, including *k*-nearest neighbor method [37] and stochastic artificial neural network [38]. The property of unknown compounds can be predicted from those compounds that are registered in a data base and have similar chemical structure.

Data-based methods are relatively simple and are applicable in almost all ADME-Tox properties. Such models require minimal computer power and have high throughputs of up to thousands or millions of molecules per hour.

Structure-based methods use three-dimensional structural information and quantum mechanics to assess the interaction potential between the small molecules studied with macromolecules, for example, enzymes or transporter proteins, that are involved in a certain ADME-Tox process [39-41]. They require 3D structures of ligands and macromolecules, and thus, require more computer power compared to the empirical models resulting in a relatively lower throughput, varying from tens to a few hundred compounds per hour.

7. Thompson TN. Optimization of metabolic stability as a goal of modern drug design. Med Res Rev 2001;21:412-49.

8. Kakkar T, Boxenbaum H, Mayersohn M. Estimation of Ki in a competitive enzyme-inhibition model: comparisons among three methods of data analysis. Drug Metab Dispos 1999;27:756-62.

9. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Del Revs 1997;23:3-25.

10. 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.

11. Delie F, Rubas W. A human colonic cell line sharing similarities with enterocytes as a model to examine oral absorption: advantages and limitations of the Caco-2 model. Crit Rev Ther Drug Carrier Syst 1997;14:221-86.

12. Yamashita S, Furubayashi T, Kataoka M, Sakane T, Sezaki H, Tokuda H. Optimized conditions for prediction of intestinal drug permeability using Caco-2 cells. Eur J Pharm Sci 2000;10:195-204.

13. Irvine JD, Takahashi L, Lockhart K, Cheong J, Tolan JW, Selick HE, Grove JR. MDCK Madin-Darby canine kidney) cells: A tool for membrane permeability screening. J Pharm Sci 1999;88:28-33.

14. Kansy M, Senner F, Gubernator K. Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. J Med Chem 1998;41: 1007-10.

15. Sugano K, Nabuchi Y, Machida M, Aso Y. Prediction of human intestinal permeability using artifical membrane permeability. Int J Pharm 2003;257:245-51.

16. Hengstler JG, Utesch D, Steinberg P, Platt KL, Diener B, Ringer M, *et al*. Cryopreserved primary hepatocytes as a constantly availble *in vitro* model for the evalution of human and animal drug metabolism and enzyme induction. Drug Metab Rev 2000;32:81-118.

17. Hakura A, Suzuki S, Satoh T. Advantage of the use of human liver S9 in the Ames test. Mutat Res 1999;438:26-36.

18. Miller VP, Stresser DM, Blanchard AP, Tuner S, Crespi CL. Fluorometric high-throughput screening for inhibitors of cytochrome P450. Ann NY Acad Sci 2000; 919:26-32.

If no structural information is available, an alternative is to use pharmacophore models, which can be built from the overlapping of known substrates of the protein. Pharmacophoric groups can be identified from structure-activity data by comparing the structure of active compounds with one another [30].

The main limitation to the widespread use of *in silico* ADME-Tox methods is their predictability [19]. The predictability of data-based models is generally limited to the chemical space that is covered by the compounds in the training set or those fairly close to them. Structure-based models, however, have an expanded chemistry space because they are based on atomic and molecular interactions between the ligand and the corresponding protein. In both cases, the use of the most diverse set of molecules available, for model development, will ensure a better prediction and broaden applicability.

## Prediction of physicochemical properties

Physicochemical properties of a drug have an important impact on pharmacokinetics; their measurement and calculation help prioritize compounds for screening, since they can be used to predict ADMET properties, and enable early decisions in drug disovery.

### pKa

The negative logarithm of the acid ionization constant (pKa) is defined as the ability of an ionizable group of an organic compound to donate a proton in an aqueous medium; ionization can affect solubility, lipophilicity, permeability and absorption of a compound.

Different algorithms were developed to calculate pKa depending on the nature of the chemical structure, including the acid-strengthening and/or base-weakening factors of the substituted aliphatic acids and bases; Taft's equation; Hammett's equation for phenols, aromatic carboxylic acid, aromatic amines, hetero-aromatic acids and bases; and the extension of Hammett's and Taft's equations to hetero-cycles. Fragment methods have proven to be very useful and are available as commercial systems [42]. Ab initio quantum mechanics calculations have been used extensively [43] as well as semi-empirical quantum mechanics [44].

Xing and Glen estimated the pKa using a novel tree structured fingerprint describing the ionizing centers [45]. Ionization models were developed using a combination of descriptors mapped onto the molecular tree constructed around the ionizable center using partial least squares with cross-validation.

### LogP and LogD

LogP is the logarithm of the partition coefficient in an octanol/water system; it refers to the neutral state of molecules and serves as a quantitative descriptor of lipophilicity. In the presence of partially dissociated compounds the ionization of a molecule is an additional factor that must be considered, since the partition then becomes pH dependent. The pH dependent distribution coefficient (logD) is related to logP through the ionization constant, pKa. Many drug molecules contain ionizable groups and hence partition across cell membranes, through pores and via active transport mechanisms that are mostly pKa dependent.

The first and still most popular commercial computer software used to calculate octanol-water partition coefficients from molecular structure is ClogP, developed by Pomona MedChem Project [20, 21].

Many computer algorithms that calculate partition coefficients have been developed since the introduction of ClogP. These methodologies can roughly be divided into (i) fragmental approaches using additive contributions of functional groups and fragments as well as their interactions such as ClogP, logKow [46] ACD/logP DB [47] and KlogP [48]; (ii) atom contribution approaches, such as AlogP98 [49] These approaches employ multiple regression equations to establish models based on a training set; (iii) topological approaches such as MlogP [50], VlogP [51]; and (iv) a neural network study [52]. Another route to logP is the direct calculation of the free energy change for transferring a solute from an aqueous to an organic solution by a thermodynamic treatment [53, 54]. These methods can usually be generalized to other two-phase systems.

### Polar surface area

The polar surface area (PSA) is commonly computed as the van der Waals surface area of all nitrogen and oxygen atoms, plus the area of the hydrogen atoms attached to them. The PSA descriptor can not distinguish between non-polar compounds or account for non-polar atom groups, thereby failing to discriminate between molecules with identical PSA but different sizes and lipophilicities. Thus, the PSA will reflect the ability of the solutes to leave the hydrogen binding environment, but not the affinity for the internal membrane or size-related effects [55].

For molecules that are able to adopt different conformations, the dynamic polar surface area (PSAd) should be a more appropriate predictor [56], To calculate the PSAd of a molecule, all three-dimensional conformations of the compounds are first constructed using molecular mechanics calculations and the PSAd is then obtained as the Boltzmann-weighted average of the van der Waal's surface areas calculated for all low-energy conformations of a compound.

Several studies have proved the PSA to be a useful descriptor of biological permeation. van der Waterbeemd and Kansy [57] were the first to correlate biological permeation with polar surface area. They found a strong correlation between brain uptake and the hydrophilic part of the calculated van der Waals surface.

Drugs with a PSA greater than 140 $Å^2$ have been found to exhibit poor intestinal absorption [28, 58-60] whereas an upper limit of 60-90 $Å^2$ has been found for blood-brain partition [59, 61].

### Lipophilicity

The lipophilicity of a drug is its tendency to prefer a lipidic, or oil-like environment to an aqueous one. It is the key physicochemical parameter linking membrane permeability (drug absorption and distribution) with the clearance (elimination) route. Although lipophilicity is a property ascribed to the drug compound, it is highly dependent on the choice of a lipidic environment [62], because behind this property lies a net of intermolecular interactions such as hydrogen bonding and dipole effects.

19. Yu H, Adedoyin A. ADME-Tox drug discovery: integration of experimental and computational technologies. Drug Discov Today 2003;8:852-61.

20. Chou JT, Jurs PC . Computer-assisted computation of partition coefficients from molecular structures using fragment constants. J Chem Inf Comput Sci 1979; 19:172-8.

21. Leo AJ, Hoekman D. Calculating log Poct) with no missing fragments the problem of estimating new interaction parameter. Perspect Drug Discov Des 2000;18:19-38.

22. Agoram B, Woltosz WS, Bolger MB. Predicting the impact of physiological and biochemical processes on oral drug bioavailability. Adv Drug Deliv Rev 2001; 50:S41-S67.

23. Stouch TR, Kenyon JR, Johnson SR, Chen XQ, Doweyko A, Li Y. *In silico* ADME/Tox: why models fail. J Comput Aided Mol Des 2003;17:83-92.

24. Van de Waterbeemd H, Rose S. In: Wermuth LG, editors. The Practice of Medicinal Chemistry 2nd; 2003. p.1367-85.

25. Buchwald P, Bodor N. Computer-aided drug design: the role of quantitative structure-property, structure-activity and structure-metabolism relationships (QSPR, QSAR, QSMR). Drug Future 2002; 15:577-88.

26. Liu R, So SS. Development of quantitative structure-property relationship models for early ADME evaluation in drug discovery. 1. Aqueous solubility. J Chem Inf Comput Sci 2001;41:1633-9.

27. Liu R, Sun H, So SS. Development of Quantitative Structure-Property Relationship Models for Early ADME Evaluation in Drug Discovery. 2. Blood-Brain Barrier Penetration. J Chem Inf Comput Sci 2001; 41:1623-32.

28. Clark DE. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. J Pharm Sci 1999;88:815-21.

29. Zhao YH, Le J, Abraham MH, Hersey A, Eddershaw PJ, Luscombe CN, *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;90:749-784.

30. Yamashita F, Hashida M. *In silico* approaches for predicting ADME properties of drugs. Drug Metab Pharmacokin 2004;19:327-38.

31. Fu XC, Chen CX, Wang GP, Liang WQ, Yu QS. Prediction of human intestinal absorption using an artificial neural network. Pharmazie 2005;60:674-6.

32. Balakin KV, Ivanenkov YA, Savchuk NP, Ivashchenko AA, Ekins S. Comprehensive computational assessment of ADME properties using mapping techniques. Curr Drug Discov Technol 2005;2:99-113.

33. Mente SR, Lombardo F. A recursive-partitioning model for blood-brain barrier permeation. J Comput Aided Mol Des 2005;19:465-81.

34. Zmuidinavicius D, Didziapetris R, Japertas P, Avdeef A, Petrauskas A. Classification structure-activity relations (C-SAR) in prediction of human intestinal absorption. J Pharm Sci 2003;92:621-33.

Most lipophilicity calculation approaches rely on fragment values, although simple methods based on molecular size and hydrogen-binding indicators for functional groups to calculate log P values have also been shown to be extremely versatile [25].

### Solubility

The solubility of drugs in water closely affects their biological activity. It is one of the important factors determining oral bioavailability. Low solubility is detrimental to good and entire oral absorption.

Several algorithms to predict aqueous solubility have been proposed. One of the simple ways to estimate a drug solubility is to use logP. It provides an estimate of the strength of the interaction of the compound with water [63]. The strong dependence of solubility on properties such as logP has formed the basis of several methods where calculated logP is augmented with additional terms [64-66].

Molecular simulations offer another route for assessing the energetic of a compound in water, Monte Carlo simulations with solute embedded in a bath of rigid water molecules to derive cohesive properties can be used to predict solubility [67], this approach performs relatively complex simulations for each different solute, thus, the computational load imposed is the main drawback; Jorgensen and Duffy have addressed this by using simulations as a way of deriving properties that can be used in a fast QSPR (quantitative structure property relationship) model.

Quantum mechanics (QM) is also used to predict solubility. An approach developed by Cramer-Truhlar [68] performs the QM calculation assuming that the compound is embedded in a continuous dielectric medium, which allows the polarization of the compound to be more accurately modeled. Klamt and co-workers [69] have produced the QM-based COSMO-RS method, which goes further by embedding both solute and solvent in a perfect conductor to calculate their polarization charge densities. Integrating over the two surfaces (solute and solvent) allows the method to calculate the chemical potential of the solute in the solvent, leading to an estimation of its solubility.

### Permeability

Cell membrane permeation is a prerequisite for drug absorption (oral, transdermal, ocular, pulmonary), distribution (across the blood-brain barrier or blood-retina barrier) and elimination by the hepatocytes in the liver, and may also undergo reabsorption through the tubular membranes of the kidney. Since the majority of drugs are administered via the oral route, the most widely studied form of biological permeation is the human intestinal absorption (HIA). Likewise, evaluation or prediction of drug permeation across the blood-brain barrier is also important, due to a large number of drugs act via the central nervous system, producing either therapeutic or adverse effects.

The main properties of a drug influencing its permeation through biological membranes are lipophilicity, hydrogen-bonding capacity, charge and size [70]. The lipophilicity is the most widely used physicochemical property to predict drug permeation in biological systems [71].

The relationship between logP and permeability is non-linear, with decreases in permeability at both low and high logP. These non-linearities are believed to be due to: (a) the limited diffusion of poorly lipophilic molecules into the phospholipid cell membrane, and (b) the preferential partitioning of highly lipophilic molecules into the phospholipid cell membrane, preventing passage [72] through the aqueous portion of the membrane [73-75].

## Prediction of ADME and related properties

### Absorption

Many factors influence the gastrointestinal tract absorption of drugs, and would fundamentally be classified into three categories: physicochemical (pKa, solubility, stability, diffusivity, lipophilicity, and salt form), physiological (gastrointestinal pH, gastric passage, small and large intestine transit time, active transport and efflux, and gut wall metabolism), and formulation factors (drug particle size and crystal form, and dosage form such as a solution, tablet, capsule, suspension, emulsion, gel, and modified release) (Figure 2) [30]. Formulation factors are usually optimized experimentally and physiological factors cannot be controlled, then prediction interests are centered on physicochemical properties of the compounds.

There are several known mechanisms of intestinal drug absorption. The major mechanism for drug uptake through the intestinal epithelium is passive diffusion driven by a concentration gradient; depending on the molecule's hydrophilicity, passive diffusion can occur through the lipid/aqueous environment of the cell membrane (trans-cellular transport) or the passage through the water-filled tight junctions formed by the fusion of lipid membranes of adjacent cells (para-cellular transport). In addition, some molecules that enter the cytoplasm of epithelial cells can be actively transported back by specific transporters to the intestinal lumen; this efflux process is mainly a function of

35. Zheng M, Liu Z, Xue C, Zhu W, Chen K, Luo X, Jiang H. Mutagenic probability estimation of chemical compounds by a novel molecular electrophilicity vector and support vector machine. Bioinformatics 2006;22:2099-106.

36. Kriegl JM, Arnhold T, Beck B, Fox T. A support vector machine approach to classify human cytochrome P450 3A4 inhibitors. J Comput Aided Mol Des 2005;19: 189-201.

37. Shen M, Xiao Y, Golbraikh A, Gombar VK, Tropsha A. Development and validation of k-nearest-neighbor QSPR models of metabolic stability of drugs candidate. J Med Chem 2003;46:3013-20.

38. Niwa T. Using general regression and probabilistic neural network to predict human intestinal absorption with topological descriptors derived from two-dimensional chemical structures. J Chem Inf Comput Sci 2003;43:113-9.

39. Jones JP, Mysinger M, Korzekwa KR. Computational models for cytochrome P450: a predictive electronic model for aromatic oxidation and hydrogen atom abstraction. Drug Metab Dispos 2002; 30:7-12.

40. Segall MD, Payne MC, Ellis SW, Tucker GT, Boyes RN. Ab Initio Molecular Modeling in the Study of Drug Metabolism. Eur J Drug Metab Pharmacokinet 1997;22:283-8.

41. Jones JP, He M, Trager WF, Rettie AE. Three-dimensional quantitative structure-activity relationship for inhibitors of cytochrome P4502C9. Drug Metab Dispos 1996;24:1-6.

42. Tsantili-Kakoulidou A, Panderi I, Csizmadia F, Darvas F. Prediction of distribution coefficient from structure. 2. Validation of Prolog D, an expert system. J Pharm Sci 1997;86:1173-9.

43. da Silva CO, da Silva EC, Nascimento MAC. Ab Initio Calculations of Absolute pKa Values in Aqueous Solution I. Carboxylic Acids. J Phys Chem A 1999;103: 11194-9.

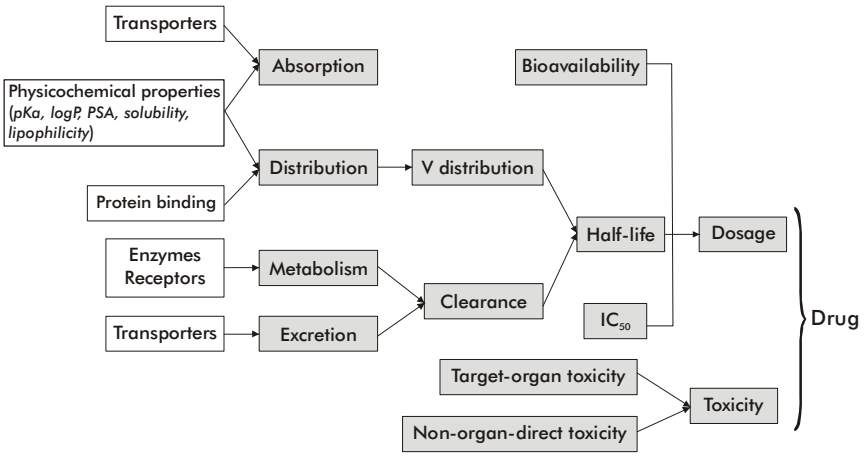


Figure 2. ADMET properties depends on many factors. Physicochemical properties are related with some of these properties such as: absorption, distribution and excretion (as a particular case of distribution). The structure of the molecule is important for its interaction with some proteins (e.g. transporters, enzymes, receptors). All these properties contribute to the pharmacokinetic of the drug candidate.

a transporter in the plasma membrane called P-glycoprotein (P-gp).

Solubility and the intestinal permeability of the drug are the most important properties that determine absorption after oral administration [76, 77]. However, their use in predicting oral absorption is a difficult task because both properties are pH dependent and they should be considered in the dynamically changing complex environment of the gastrointestinal tract.

Almost all of the computational approaches currently used to predict absorption are based on the assumption that absorption is passive, and can be predicted from molecular descriptors of the compound. Perhaps the most widely used model for the prediction of passive intestinal absorption is the Rule-of-five model introduced by Lipinski [9]. The Rule-of-five considers that a molecule will be poorly absorbed or poorly soluble when the following cut-offs are exceded: logP > 5, molecular weight > 500, the number of hydrogen bond (H-bond) donors (counted as hydrogens attached to N or O atoms) > 5, or the number of H-bond acceptors (approximated as the number of N and O atoms) > 10, and to have particularly poor absorption or solubility if any two of these bounds are exceeded. The Rule-of-five model has the advantages of being simple, easy to interpret, and fast to compute [78]. However it has recently been described that many natural products are exceptions of the Rule-of-five, and might also show good oral bioavailability [79], as well as others compounds with molecular weight (MW) > 500 but with reduced molecular flexibility and constrained polar surface area [60, 79].

In the modeling of the intestinal absorption of molecules, the PSA is one of the most widely used descriptors. Several reports are available showing an experimental correlation of PSA with the apparent permeability through a Caco-2 monolayer [80] or Caco-2 cells and rat ileum [58]. The PSAd values take into account the shape and flexibility of a drug, while it has also been shown that the PSA calculated for a representative single conformer performs as well as the PSAd used to predict permeability [28].

Wessel and co-workers have reported a neural network model to predict percent human intestinal absorption (%HIA) [81]. The descriptors they used were topological, electronic, geometric, charged-partial surface area (CPSA) and other related ones. These methods, however, are not applicable to a large number of compounds, because estimations of PSA and CPSA require conformational analysis and molecular orbital calculations, which are computationally impracticable for a large set of compounds.

Others also use PSA to predict oral absorption [31] developing an artificial neural network model to predict %HIA of compounds from their molecular structure parameters. These parameters are the PSA, the fraction of polar molecular surface area (FPSA, polar molecular surface area/ molecular surface area), the sum of the net atomic charges of oxygen atoms (Q(O)), the sum of the net atomic charges of nitrogen atoms with net negative atomic charges (Q(N)), the sum of the net atomic charges of hydrogen atoms attached to oxygen or nitrogen atoms (Q(H)), and the number of carboxyls (nCOOH).

Other approach to predict Caco-2 cell permeability was developed by Fujiwara and co-workers [82]. They calculated molecular descriptors of structurally diverse compounds with a semi-empirical molecular orbital calculation method, and then applied an artificial neural network to the multivariate analysis between molecular descriptors and Caco-2 cell permeability. This approach was compared with a multiple linear regression with respect to the predictability of Caco-2 cell permeability and revealed that the neural network model had a fairly good predictability as far as Caco-2 cell permeability is concerned, and better than the simple and quadratic regression model.

A methodology combining a genetic algorithm search with neural network analysis applied to the modeling of Caco-2 cell apparent permeability was developed by Di Fenza and co-workers [83]. Several molecular descriptors of the compounds were calculated and the optimal subsets were selected using a genetic algorithm. The selected descriptors were shown to possess physico-chemical connotations which are in excellent accordance with the well-known relevant molecular properties involved in the cellular membrane permeation phenomenon: hydrophilicity, hydrogen bonding propensity, hydrophobicity and molecular size. The predictive ability of the models, although rather good for a preliminary study, is somewhat affected by the poor precision of the experimental Caco-2 measurements. The generalization ability of one model was checked on an external test set. The result obtained is of interesting practical application and stresses that a successful model construction is strictly dependent on the structural space representation of the data set used for model development.

Linear discriminant analysis (LDA) has been used to obtain quantitative models that discriminate higher absorption compounds from those with a moderate to poorer absorption [84]. The models were based on a data set of measured Caco-2 cell permeability, consisting of 157 structurally diverse compounds. The best LDA model has an accuracy of 90.58% and 84.21% for training and test set, respectively. In a virtual screening of 241 drugs with the reported values of the percentage of human intestinal absorption (%HIA), the percentage of good correlation was greater than 81%. In addition, multiple linear regression models were developed to predict Caco-2 permeability with determination coefficients of 0.71 and 0.72. These results suggest that the proposed method is a good tool for studying the oral absorption of drug candidates [84].

Two dimension (2D) descriptors have also been used for human intestinal absorption prediction. Yamashita and co-workers [85] predicted Caco-2 cell permeability from 2D topological descriptors, which was optimized by a genetic algorithm combined with partial least squares. The predictability of this model was comparable to that of the artificial neural network model where the same permeability data were predicted from quantum chemical descriptors [82].

A rapid and reliable method to predict the %HIA of compounds based on their 2D descriptors was developed by Niwa [38]. The %HIA values were modeled using a general regression neural network and a probabilistic neural network, variants of normalized radial basis function networks. Both networks performed well to model the %HIA values.

44. Citra MJ. Estimating the pKa of phenols, carboxylic acids and alcohols from semiempirical quantum chemical methods. Chemosphere 1999;38:191-206.

45. Xing L, Glen RC. Novel methods for the prediction of logP, pKa, and logD. J Chem Inf Comput Sci 2002;42:796-805.

46. Meylan WM, Howard, PH. Atom/fragment contribution method for estimating octanol-water partition coefûcients. J Pharm Sci 1995;84:83-92.

47. Spessard GO. ACD Labs/LogP dB 3.5 and ChemSketch 3.5. J Chem Inf Comput Sci 1998;38:1250-3.

48. Klopman G, Li J-Y, Wang S, Dimayuga M. Computer automated log P calculations based on an extended group contribution approach. J Chem Inf Comput Sci 1994; 34:752-81.

49. Ghose AK, Viswanadhan VN, Wendoloski JJ. Prediction of hydrophobic (lipophilic) properties of small organic molecules using fragmental methods: an analysis of ALOGP and CLOGP methods. J Phys Chem A 1998;102:3762–3772.

50. Moriguchi I, Hirono S, Liu Q, Nakagome I, Matsushita Y. Simple method of calculating octanol/water partition coefficient. Chem Pharm Bull 1992;40: 127-30.

51. Gombar VK, Enslein K. Assessment of n-octanol/water partition coefficient: when is the assessment reliable?. J Chem Inf Comput Sci 1996;36:1127-34.

52. Huuskonen JJ, Livingstone DJ, Tetko IV. Neural network modeling for estimation of partition coefûcient based on atom-type electrotopological state indices. J Chem Inf Comput Sci 2000;40:947-55.

53. Reynolds CH . Estimating lipophilicity using the GB/SA continuum solvation model: A direct method for computing partition coefficients. J Chem Inf Comput Sci 1995;35:738-42.

54. Ruelle P. Universal model based on the mobile order and disorder theory for predicting liphphilicity and partition coefficients in all mutually immiscible two-phase liquid systems. J Chem Inf Comput Sci 2000;40:681-700.

55. Mälkiä A, Murtomäki L, Urtti A, Kontturi K. Drug permeation in biomembranes *in vitro* and *in silico* prediction and influence of physicochemical properties. Eur J Pharm Sci 2004;23:13-47.

56. Palm K, Luthman K, Ungell A-L, Strandlund G, Artursson P. Correlation of drug absorption with molecular surface properties. J Pharm Sci 1996;85:32-9.

57. van de Waterbeemd H, Kansy M. Hydrogen-bonding capacity and brain penetration. Chimia 1992;46:299-303.

58. Palm K, Stenberg P, Luthman K, Artursson P. Polar molecular surface properties predict the intestinal absorption of drugs in human. Pharm Res 1997;14: 568-71.

59. Kelder J, Grootenhuis PDJ, Bayada DM, Delbressine LPC, Ploemen JP. Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. Pharm Res 1999;16: 1514-9.

60. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 2002;45:2615-23.

Hou and co-workers developed models to predict human intestinal absorption by using the genetic function approximation technique, and for a training set of 455 compounds [86]. The best prediction model contains four molecular descriptors: topological polar surface area, the predicted distribution coefficient at pH = 6.5, the number of violations of the Lipinski's rule-of-five, and the square of the number of hydrogen-bond donors. The high quality of the classification model was validated by the satisfactory predictions on the training set (correctly identifying 95.9% of the compounds in the poor-absorption class and 96.1% of the compounds in the good-absorption class) and on the test set (correctly identifying 100% of the compounds in the poor-absorption class and 96.8% of the compounds in the good-absorption class).

In another approach, Hou and co-workers studied the performance of a support vector machine (SVM) to classify compounds with high or low fractional absorption (%FA > 30% or %FA < or = 30%) [87]. SVM classification models have been generated to investigate the impact of specific molecular properties on % FA. Among these important molecule descriptors, the topological polar surface area (TPSA) and predicted apparent octanol-water distribution coefficient at pH 6.5 (logD6.5) show the best classification performance. The analyzed data set consists of 578 structural diverse drug-like molecules, which have been divided into a 480-molecule training set and a 98-molecule test set. The best SVM classifier can give satisfactory predictions for the training set (97.8% for the poor-absorption class and 94.5% for the good-absorption class). Moreover, 100% of the poor-absorption class and 97.8% of the good-absorption class in the external test set could be correctly classified. This illustrates that SVMs combined with simple molecular descriptors can provide an extremely reliable assessment of intestinal absorption in an early *in silico* filtering process.

Ito and co-workers [88] have developed a pharmacokinetic model for drug absorption that includes metabolism by CYP3A4 inside the epithelial cell, P-gp-mediated model efflux into the lumen, intracellular diffusion from luminal side to basal side, and subsequent permeation through the basal membrane. They demonstrated that the fraction absorbed was elevated, by the simultaneous inhibition of both CYP3A4 and P-gp.

A set of well-defined structural elements required for interaction with P-gp has been derived from the analysis of a set of known P-gp substrates [89-91]. The key recognition elements in this model are two or three electron-donor groups with a fixed spatial separation. The inhibition of P-gp would increase the intestinal absorption of P-gp substrates [92].

The MolSurf program has been used to generate descriptors to build a PLS model to predict P-gp-associated ATPase activity [93]. This model identified the main contributing descriptors for predicting ATPase activity as the size of the molecule surface, polarizability and hydrogen-bonding potential.

Swaan et al [94] examined the structure-affinity relationship for the small intestinal oligopeptide carrier (PepT1) using CoMFA, the model obtained showed a high correlation between carrier permeability and steric (73% contribution) and electrostatic (23% contribution) molecular fields with a cross-validate r2 of 0.754, besides, CoMFA has been used in modeling the intestinal bile acid carrier [95], and P-gp [96, 97].

A P-gp pharmacophore model consisting of two hydrophobic points, three hydrogen-bond-acceptor points and one donor point was reported [98]. A three-dimensional QSAR P-gp model was generated using the Catalyst program [99] and this model made it possible to qualitatively rank predicting IC50 values for P-gp inhibitors.

Crivori and co-workers developed several models to predict substrates and inhibitors of P-gp. A method for discriminating P-gp substrates and non-substrates has been set up based on calculated molecular descriptors and multivariate analysis [100]. The drugs of the training set were previously classified as P-gp substrates or non-substrates on the basis of the efflux ratio from Caco-2 permeability measurements. The descriptors were calculated with the Volsurf program and were correlated to the experimental classes using partial least squares discriminant analysis. The model was able to correctly predict the behavior of 72% of an external set of 272 proprietary compounds.

On the basis of the P-gp inhibition data, a partial least squares discriminant analysis using GRIND-pharmacophore-based descriptors was performed to model P-gp substrates having poor or no inhibitory activity versus inhibitors having no evidence of significant transport [100]. The model was able to discriminate between 69 substrates and 56 inhibitors taken from the literature with an average accuracy of 82%. The model also allowed the identification of some key molecular features that differentiate a substrate from an inhibitor, which should be taken into consideration in the design of new candidate drugs.

A robust predictive pharmacophore model was targeted in a supervised analysis of three-dimensional (3D) pharmacophores from 163 published compounds to differentiate non-substrates from substrates of Pgp [101]. A comprehensive set of pharmacophores has been generated from conformers of whole molecules of both, substrates and non-substrates of P-glycoprotein. Four-point 3D pharmacophores were employed to increase the amount of shape information and resolution, including the ability to distinguish chirality. A novel algorithm of the pharmacophore-specific t-statistic was applied to the actual structure-activity data and 400 sets of artificial data (sampled by decorrelating the structure and Pgp efflux activity). A simple classification tree using nine distinct pharmacophores was constructed to distinguish non-substrates from substrates of Pgp. An overall accuracy of 87.7% was achieved for the training set and 87.6% for the external independent test set. Furthermore, each of nine pharmacophores can be independently used as an accurate marker for potential Pgp substrates [101].

Software packages are available commercially for predicting human intestinal fraction absorbtion based on estimates of solubility and intestinal permeability (Table 1).

GastroPlus simulates gastrointestinal absorption and pharmacokinetics for drugs administered orally or intavenously in humans and animals. Oral absorp-

61. Van de Waterbeemd H, Camenisch G, Folkers G, Chretien JR, Raevsky OA. Estimation of blood-brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. J Drug Target 1998;6:151-65.

62. Kaliszan R . Quantitative structure-retention relationships. Anal Chem 1992: 619A-631A.

63. Sanghvi T, Jain N , Yang G , Yalkowsky SH. Estimation of aqueous solubility by the general solubility equation (GSE) the easy way. QSAR & Combinatorial Science 2003;22:258-62.

64. Butina D, Gola JMR. Modeling aqueous solubility. J Chem Inf Comput Sci 2003;43: 837-41.

65. Cheng A, Merz KM. Prediction of aqueous solubility of a diverse set of compounds using quantitative structure-property relationship. J Med Chem 2003; 46:3572-80.

66. Delaney JS. ESOL: Estimating aqueous solubility directly from molecular structure. J Chem Inf Comput Sci 2004;44: 1000-5.

67. Jorgensen WL, Duffy EM. Prediction of drug solubility from Monte Carlo simulations. Bioorg Med Chem Lett 2000; 10:1155-8.

68. Cramer CJ, Truhlar DG . Continuum solvation models: Classical and quantum mechanical implementations. In: Lipkowitz KB, Boyd DB, editors. Reviews in Computational Chemistry; 1995. p.1-72.

69. Klamt A. Prediction of aqueous solubility of drugs and pesticides with COSMO-RS. J Comput Chem 2002;23:275-81.

70. Camenisch G, Folkers G, van de Waterbeemd H. Review of theoretical passive drug absorption models: historical background, recent developments and limitations. Pharm Acta Helv 1996; 71:309-27.

71. Testa B, Carrupt P-A, Gaillard P, Billois F, Weber P. Lipophilicity in molecular modelling. Pharm Res 1996;13:335-43.

72. Gobas FA, Lahittette JM, Garofolo G, Shiu W, MacKay D. A novel method for measuring membrane-water partition coefficients of hydrophobic organic chemicals: comparison with 1-octanol-water partitioning. J Pharm Sci 1988;77: 265-72.

73. Kararli TT. Gastrointestinal absorption of drugs. Cri. Rev Ther Drug Carrier Syst 1989;6:39-86.

74. Hansch C, Clayton JM. Lipophilic character and biological activity of drugs. II. Parabolic case. J Pharm Sci 1973;62: 1-21.

75. Martin YC. A practitioner's perspective of the role of quantitative structure-activity analysis in medicinal chemistry. J Med Chem 1981;24:229–237.

76. Sinko PJ, Leesman GD, Amidon GL. Predicting fraction dose absorbed in humans using a macroscopic mass balance approach. Pharm Res 1991;8:979-88.

77. Oh DM, Curl RL, Amidon GL. Estimating the fraction dose absorbed from suspensions of poorly soluble compounds in humans: a mathematical model. Pharm Res 1993;10:264-70.

Table 1. Programs for *in silico* prediction of ADME-Tox properties

| Program | Developer | Purpose and Function | Models used | Instrumentation |
|---|---|---|---|---|
| GlastroPlus | Simulation-plus, Inc. www.simulation-plus.com | Simulations and predictions of the gastrointestinal dissolution, transit, absorption, bioavailability, and pharmacodynamics. Prediction of the first-passage effect in the gut and liver and plasma concentration-time profiles. | Physiologically-based mechanistic advanced compartmental absorption and transit model | Gastrointestinal simulation linked to pharmacokinetics and pharmacodynamic models |
| iDEA | LionBioscience, Inc. http://www.lionbioscience.com/solutions/idea | Absorption Module predicts the fraction dose absorbed over time, mass absorbed, soluble mass, insoluble mass, absorption rate, and intestinal drug concentration. Metabolism Module predicts the extent of first pass metabolism for a compound. | A knowledge based model relying on *in vitro* data and clinical data of a proprietary database of chemical structures | Uses chemical structure to predict various absorption properties. Absorption Model uses *in vitro* data to construct a physiological model. Metabolism Model is linked to the Absorption Module for inputs and predicts bioavailability from metabolic turnover and protein binding data |
| COMPACT (computer-optimized molecular parametric analysis of chemical toxicity) | University of Surrey, Guildford, UK | Identification of potential carcinogenicity and toxicity mediated by one or more cytochromes P450. | Enzyme modeling systems based on theoretical and mechanistic considerations | Computes the shape and molecular orbital energy levels of a chemical structure and evaluates whether it can interact with the active site of cytochrome P450 I or to the binding site of the Ah receptor and thereby induce cancer |
| Camitro | Camitro Corporation | Predicts aqueous solubility, blood-brain barrier partitioning, human intestinal absorption and cytochrome P450 (CYP3A4) metabolism and CYP (CYP2D6 and 2C9) inhibitory potential. | Enzyme modeling systems based on theoretical and mechanistic considerations | Uses surface and electronic properties of the molecule taking into account theoretical energy differences by reaction/diffusion. Metabolism models combined empirical/quantum chemical approach to predict enzyme-substrate binding affinities, metabolic sites, and relative rates of metabolism |
| MetabolExpert | CompuDrug Chemistry Ltd www.compdrug.com | Generates metabolites with quickly identifying sites on the molecule where metabolic transformation may occur. | A knowledge base of structure-metabolism rules with open architecture | Rules based on examples from the literature and on the basis of possible sites and restrictions from the compound under study |
| METEOR | Lhasa, Ltd www.chem.leeds.ac.uk/luk/meteor | Prediction of the metabolic fate of a query chemical structure. | A knowledge base of structure-metabolism rules together with a reasoning engine | The only information needed is the molecular structure of the chemical The reasoning model takes into account the lipophilicity (log P estimate) and the most likely metabolites generated |
| META | Case Western Reserved University, Cleveland, OH, USA | Predict the sites of potential enzymatic attack and the nature of the chemical formed by such metabolic transformations | empirically based expert system | Uses dictionaries of biotransformation operators. Biotransformations are based on recognition of key functional groups. Different biotransformations receive different priorities |
| MultiCASE | MultiCASE Inc. www.multicase.com | Use databases for carcinogenicity, mutation, teratogenicity, biodegradation, endocrine disruption | Knowledge-based systems | Divides the molecule into various molecular fragments, then creates organized dictionaries of these fragments and develops QSAR correlations to determine which fragments are relevant to model toxicity |
| DEREK | Lhasa, Ltd www.chem.leeds.ac.uk/luk/derek | Predicts genotoxicity, mutagenicity, carcinogenicity, skin sensitization. It highlights potential toxicological hazards covering a wide range of endpoints from irritancy to hepatotoxicity | A knowledge base that contains alerts describing structure-toxicity relationships | Uses structure-based alerts that define toxophores. Alerts cover a wide range of toxicological end points, including carcinogenicity, mutagenicity, and skin sensitization |
| TOPKAT | Accelrys www.accelrys.com | Predicts toxicity endpoints based on chemical structure, including rodent carcinogenicity, Ames mutagenicity, rat oral LD50, rat chronic lowest-observable adverse effect level, developmental toxicity potential and skin sensitization | Statistically based system that consists of a suite of QSAR models | Robust and cross-validated QSAR models, based upon two-dimensional molecular, electronic and spatial descriptors, for assessing various measures of toxicity |
| OncoLogic | EPA's Office of Pollution Prevention and Toxics (OPPT) and LogiChem, Inc http://www.epa.gov/oppt/newchems/tools/oncologic.htm | Predicts the carcinogenic potential of chemicals | based on knowledge rules | The program applies SAR analysis to predict the potential cancer-causing effects of a chemical and applies the knowledge gained from studies of how chemicals cause cancer in animals and humans |

Table 1. Programs for *in silico* prediction of ADME-Tox properties

| Program | Developer | Purpose and Function | Models used | Instrumentation |
|---|---|---|---|---|
| HazardExpert | CompuDrug Chemistry Ltd www.compudrug.com | Prediction of the compound's toxicity class: oncogenicity, mutagenicity, teratogenicity, membrane irritation, sensitivity, immunotoxicity, neurotoxicity. | rule-based system | An artificial neural network based approach using atomic fragmental descriptors. Predicts the toxicity of organic compounds based on toxic fragments. |
| VirtualToxLab | Biographics Laboratory 3R http://www.biograf.ch/index.php?id=projects&subid=virtualtoxlab | Provides access to a series of mQSAR models for the estimation of the toxic potential of chemical compounds towards selected human target receptors (e.g. AhR, AR, ERα/β, GR, PPARγ, TRα/β) and enzymes (e.g. CYP 3A4). | multidimensional quantitative structure-activity relationship model | mQSAR platform -featuring the Quasar and Raptor technologies- augmented by Yeti/AutoDock and Symposar for generating the pharmacophore hypotheses. The VirtualToxLab runs on Macintosh, Linux and Unix platforms. The interface has been written using the Java programming language to allow for an easy and platform-independent access. |

tion simulation is based on an advanced version [22] of the compartmental absorption-transit model [102] and provides estimates of the fraction of the dose absorbed. Permeability in each compartment is scaled according to the pH of that compartment, the logP and the pKa values of the drug. QMPRPlus must be used for pure *in silico* predictions, which takes an input file of multiple structures and generates estimates for lipophilicity (logP), effective permeability, apparent permeability, diffusivity, and water solubility; the estimates of QMPRPlus are derived from correlation models using a variety of data from human and *in vitro* studies together with primary molecular descriptors of chemical structures.

iDEA simulates human physiology and accounts for regional variations in intestinal permeability, solubility, surface area and fluid movement. The system is based on the STELLA (Structural Thinking Experimental Learning Laboratory with Animation) simulation software, a physiologically based absorption model defining each intestinal segment as a separate compartment. It is used for the description of fluid movement in the gastrointestinal tract with a calculation of drug absorption in each intestinal segment over time [103, 104]. The summation of the flux calculations in each segment gives the total absorption rate. The absorption model is coupled with a physiological metabolism model, which provides estimates for the rate and extent of first pass metabolism in humans; the combined system allows the prediction of bioavailability for a compound from *in vitro* data.

## Distribution

Once a compound is absorbed, it must be distributed in the body to reach its target. The distribution will depend on the structural and physicochemical properties of the compound (Figure 2), Most drugs exhibit a non-uniform distribution in the body with variations that are largely determined by their ability to pass through membranes and their lipid/water solubility.

The initial interaction of the compounds will be with plasma proteins, these proteins exert a large influence on the distribution process because drugs can bind to a variety of them such as albumin (acidic drugs), alpha1-acid glycoprotein (basic drugs) lipoproteins (neutral and basic drugs), erythrocytes and alpha, beta, gamma-globulins [105]. Considering the high concentration of albumin, the drug-free concentration can be effectively reduced for drugs strongly bound to plasma proteins, although the affinity of drugs for plasma proteins is usually lower than for the receptor or enzyme targets.

Drug-protein complexes in plasma also serve as drug reservoirs. The effective concentration to be targeted and the potential side effects influence the amount of binding to plasma proteins that can still be tolerated and how precisely this parameter has to be tuned for a new drug entity [105]. Tissue distribution is an important determinant of the pharmacokinetic profile of a drug; an its understanding may also help in predicting the pharmacodynamic effects of a drug in specific tissues.

Drug distribution can be estimated using tissue: plasma ratios or the volume of a steady state distribution (Vss). There are several methods available to predict tissue distribution using either tissue: plasma ratios or the volume of distribution at a steady state (Vss). All methods are based on the assumption that there is a passive diffusion between tissue compartments, even when for some drugs active influx or efflux can play an important role in determining the volume of distribution at a steady state (Vss).

The identification of the tissue: plasma partition coefficients (Kp) needed for an initial prediction of the volume of steady state distribution (Vss) of a drug in humans was studied by Bjorkman [106]. There were excellent linear correlations between Vss calculated by means of only two Kp values and the originally calculated Vss. Thus, the initial estimation of the Vss of a new drug can normally be based on only two Kp values, those of muscle and fat. The muscle Kp can be used to represent all lean tissues, including the residual "carcass", and fat Kp can be used for the distribution of basic drugs to the lungs.

Physiological information on tissue composition (lipid/water/protein fraction), the blood composition (lipid/water/protein) and the blood flowing into the tissues are often used to develop a partitioning model. Poulin and co-workers [107, 108] developed tissue composition-based equations for calculating tissue-plasma partition coefficients based on the assumption that each tissue and plasma is a mixture of lipids, water, and plasma proteins in which the drug can be homogeneously distributed. They found that 80% of all predicted values were within a factor of two of the corresponding experimental values. In another approach,

78. Egan WJ, Lauri G. Prediction of intestinal permeability. Adv Drug Deliv Rev 2002;54:273-89.

79. Keller TH, Pichota A, Yin Z. A practical view of 'druggability'. Curr Opin Chem Biol 2006;10:357-61.

80. Van de Waterbeemd H, Camenisch G, Folkers G, Raevsky OA. Estimation of Caco-2 Cell Permeability Using Calculated Molecular Descriptors. Quant Struct-Act Relat 1996;15:480-90.

81. Wessel MD, Jurs PC, Tolan JW, Muskal SM. Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure. J Chem Inf Comput Sci 1998;38:726-35.

82. Fujiwara S, Yamashita F, Hashida M. Prediction of Caco-2 cell permeability using a combination of MO-calculation and neural network. Int J Pharm 2002; 237:95-105.

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

84. Castillo-Garit JA, Marrero-Ponce Y, Torrens F, Garcia-Domenech R. Estimation of ADME Properties in Drug Discovery: Predicting Caco-2 Cell Permeability Using Atom-Based Stochastic and Non-stochastic Linear Indices. J Pharm Sci 2008; 97:1946-76.

85. Yamashita F, Wanchana S, Hashida M. Quantitative structure / property relationship analysis of Caco-2 permeability using genetic algorithm-based partial least squares method. J Med Chem 2002;91:2230-9.

86. Hou T, Wang J, Zhang W, Xu X. ADME evaluation in drug discovery. 7. Prediction of oral absorption by correlation and classification. J Chem Inf Model 2007; 47:208-18.

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

88. Ito K, Kusuhara H, Sugiyama Y. Effects of intestinal CYP3A4 and P-glycoprotein on oral drug absorption-theoretical approach. Pharm Res 1999;16:225-31.

89. Seelig A. A general pattern for substrate recognition by P-glycoprotein. Eur J Biochem 1998;251:252-61.

Poulin and Theil [109] illustrated physiologically based pharmacokinetic models to estimate "*a priori*" the overall plasma and tissue kinetic beha-vior under "*in vivo*" conditions. Tissue distribution of two lipophilic bases and one neutral, more hydrophi-lic, drug were predicted by this approach. The results indicated that most of the simulated concentration-time profiles of plasma and 10 tissues were in reaso-nable agreement with the corresponding experimental data determined *in vivo*.

Such approaches have been evaluated in the pharmaceutical industry and have been shown to be of considerable value in early drug discovery for predicting Vss and tissue distribution predicting within 0.5-2.0-fold of the Vss value determined *in vivo*. This could be improved by including additional considerations for certain types of compounds [110].

Plasma protein binding is also taken into consideration, [110] using three simple measurements to predict the distribution volume of a new compound: human plasma protein binding, experimental logD and pKa. The method relies on a correlation derived between the unbound fraction in tissues, calculated from the human volume of distribution data and human plasma protein binding using the Oie-Tozer equation [111], and a composite of physicochemical properties. This method is as accurate as other reported methods based on animal pharmacokinetic data, using a similar set of compounds, and ranges between 1.62 and 2.20 as mean-fold error.

Another approach was developed to predict the human serum albumin binding (logK(HSA)) of neutral molecules using the set of 5 COSMO-RS sigma-moments as descriptors [112]. This model was built on a data set of 92 compounds and achieved an r2 of 0.67 and an rms error of 0.33 log units. The model was validated by leave-one-out cross-validation tests, which resulted in q2 = 0.63 and a qms error of 0.35 for the logK(HSA) model.

The QSAR derived from the analysis of the relationship between physicochemical data/properties/ structures and experimental data on Vss or tissue: plasma ratios from *in vivo* studies with specific groups of compounds has been used in Vss predictions. A QSAR tool was developed to predict human and rat Vss and used three different statistical methodologies: Bayesian neural networks, classification and regression trees, and partial least squares [113]. Results indicated that human and rat models could be very useful in the early stages of the drug discovery process.

QSPR techniques for the prediction of volume of distribution have also been used [114]. Structural descriptors consisted of partitioning, quantum mechanics, molecular mechanics, and connectivity parameters and genetic algorithm or step-wise regression analyses were used in the selection of the variable and model development, validating the models by a leave-many-out procedure. QSPR analyses resulted in a number of significant models for acidic and alkaline drugs separately, and for all drugs; although separate QSPR models for acidic and alkaline drugs resulted in lower prediction errors than models for all drugs, an external validation study showed a limited applicability for the equation obtained for acidic drugs.

### *Blood-brain barrier penetration*

The blood-brain barrier (BBB) is an important element in the regulation of the internal environment of the brain. Drugs that act in the CNS need to cross the BBB, in contrast, drugs with a peripheral target may not cross the BBB to avoid CNS side effects.

Tight junctions between endothelial cells as well as the lack of an aqueous pathway between cells restrict the movement of polar molecules across the cerebral endothelium [115]. In addition to specific transport mechanisms (P-glycoprotein or receptor-mediated transport, peptide transporters, and other transport systems such as GLUT-1, system L1, and system ASC) passive diffusion is one of the most important ways to penetrate the barrier.

'Rule-of-five'-like recommendations regarding the molecular parameters that contribute to the ability of molecules to cross the BBB have been made to aid BBB-penetration predictions [61]; for example, molecules with molecular mass of < 450 Da or with PSA < 100 $A^2$ are considered more likely to penetrate the BBB.

Early predictions logBB (logarithm value of brain to plasma concentration ratio) involved classical QSAR approaches using various physicochemical parameters such as the octanol-water partition coefficient (logP) [116], molecular size descriptors [117, 118] and solvation parameters [119, 120]. A combination of molecular descriptors was used to predict log BB using lipophilicity, polarity, polarizability, and hydrogen-bonding parameters and partial least-squares statistics [121], and it was demonstrated that topological and constitutional descriptors used in partial least-squares correlate well with experimental logBB data [122]. Although the QSPR methodology was found to be useful in the accurate prediction of logBB data for a relatively small set of compounds, other methods combined QSPR and PSA with molecular volumes to yield a function used to calculate logBB giving an acceptable accuracy [57].

Abraham and co-workers [119, 120] developed two successful linear models. The descriptors they used include logP, excess molar refraction, dipolarity/ polarizability, hydrogen bonding acidity, hydrogen-bonding alkalinity, and McGowan molecular volume. A potential problem in their models is that the descriptors are not easy to calculate for structurally diverse drug candidates. Afterwards, more studies were published [28, 59, 121-125] using molecular descriptors that are easier to calculate such as molecular volume, surface area, shape, topological indices, logP, etc. Most of these studies use some conformationdependent and/ or experimentally determined descriptors and therefore do not conform to the desirable features stated above.

There is a model using PSA as the only descriptor [59], calculating the PSAd from all low-energy conformations of a particular compound. The results showed an excellent correlation with the experimental logBB data [59], for the protocol, however, it includes an extensive conformational analysis for each molecule, which clearly prevents its application in a virtual high-throughput screening. A similar model was also developed by Clark [28], but these models cannot distinguish the difference in BBB penetration of hydrocarbon

90. Seelig A, Landwojtowicz E. Structure-activity relationship of P-glycoprotein substrates and modifiers. Eur J Pharm Sci 2000;12:31-40.

91. Seelig A, Blatter XL, Wohnsland F. Substrate recognition by p-glycoprotein and the multidrug resistance-associated protein MRP1: a comparison. Int J Clin Pharmacol Ther 2000;38:111-21.

92. Li AP. Screening for human ADME/Tox drug properties in drug discovery. Drug Discov Today 2001;6:357-66.

93. Österberg Th, Norinder U. Theoretical calculation and prediction of P-glycoprotein-interacting drugs using MolSurf parametrization and PLS statistics. Eur J Pharm Sci 2000;10:295-303.

94. Swaan PW, Koops BC, Moret EE, Tukker JJ. Mapping the binding site of the small intestinal peptide carrier (PepT1) using comparative molecular field analysis. Receptors Channels 1998;6:189-200.

95. Swaan PW, Szoka FC Jr, Oie S. Molecular modeling of the intestinal bile acid carrier: a comparative molecular field analysis study. J Comput Aided Mol Des 1997;11:581-8.

96. Pajeva I, Wiese M. Molecular modeling of phenothiazines and related drugs as multidrug resistance modifiers: a comparative molecular field analysis study. J Med Chem 1998;41:1815-26.

97. Yates CR, Chang C, Kearbey JD, Yasuda K, Schuetz EG, Miller DD, Dalton JT, Swaan PW. Structural determinants of P-glycoprotein-mediated transport of glucocorticoids. Pharm Res 2003;20: 1794-803.

98. Pajeva IK, Wiese M. Pharmacophore model of drugs involved in P-glycoprotein multidrug resistance: Explanation of structural variety (Hypothesis). J Med Chem 2002;45:5671-86.

99. Ekins S, Kim RB, Leake BF, Dantzig AH, Schuetz EG, Lan L-B, *et al*. Three-dimensional quantitative structure-activity relationships of inhibitors of P-glycoprotein. Mol Pharmacol 2002;61:964-73.

100. Crivori P, Reinach B, Pezzetta D, Poggesi I. Computational models for identifying potential P-glycoprotein substrates and inhibitors. Mol Pharm 2006;3:33-44.

101. Li WX, Li L, Eksterowicz J, Ling XB, Cardozo M. Significance analysis and multiple pharmacophore models for differentiating P-glycoprotein substrates. J Chem Inf Model 2007;47:2429-38.

102. Yu LX, Amidon GL. A compartamental absorption and transit model for estimating oral drug absorption. Int J Pharm 1999;186:119-25.

103. Grass GM. Simulation models to predict oral drug absorption from *in vitro* data. Adv Drug Deliv Rev 1997;23: 2512-572.

104. Norris DA, Leesman GD, Sinko PJ, Grass GM. Development of predictive pharmacokinetic simulation models for drug discovery. J Controlled Release 2000;65:55-62.

105. Kratochwil NA, Huber W, Muller F, Kansy M, Gerber PR. Predicting plasma protein binding of drugs: a new approach. Biochem Pharmacol 2002;64:1355-74.

compounds, because by definition PSA is the molecular surface area contributed by nitrogen and oxygen atoms only. In an effort to account for hydrophobic contributions, Clark introduced logP as an additional descriptor [28], demonstrating that static PSA combined with ClogP [20, 21] is a good estimate of logBB.

Efforts have been made in seeking a better hydrophobicity descriptor for drug transport properties. Based on the consideration that to be transported across the lipid bilayer, a molecule must pass through the outer hydrated polar part of the bilayer as well as through the much more hydrophobic membrane inside, one would naturally think that an ideal parameter should be similar or related to logP. Young et al. investigated the performance of the $logP_{cyclohexane\text{-}water}$ partition coefficient and found that the difference ($\Delta$logP) between $logP_{oct}$ and $logP_{cyclohexane\text{-}water}$ correlates better with the logBB of 20 histamine H2 receptor antagonists [126], and other authors found that the brain permeability of a series of structurally diverse histamine H1 receptor antagonists was better explained by $logD_{oct}$ than by $\Delta$logP [127].

Combinations of molecular descriptors such as lipophilicity, polar and hydrophobic surface areas, structural parameters and net charge at physiological pH, have been used for the prediction of passive membrane transport. Computational PLS regression models using these descriptors were developed to predict CNS penetration of drug-like organic molecules [128]. For modeling, a dataset of 77 structurally diverse compounds was used with reported steady-state rat brain to plasma ratios (BPR). Information on steady-state cerebrospinal fluid distribution (CSF to plasma ratio or CSFPR) was available for 37 of these compounds. They were CNS active and were therefore assumed to penetrate the blood-brain barrier and/or the blood-fluid barrier. Using these PLS models, the dataset could be described accurately (r(2) = 0.78, StErrorEst = 0.30 and r(2) = 0.75, StErrorEst = 0.28 for BPR and CSFPR, respectively). The present models provide a cost-effective and efficient strategy to guide synthetic efforts in medicinal chemistry at an early stage of the drug discovery and development process.

A model for the prediction of blood-brain partitioning (logBB) was developed using the set of 5 COSMO-RS sigma-moments as descriptors [112]. Sigma-moments were obtained from quantum chemical calculations using the continuum solvation model COSMO and a subsequent statistical decomposition of the resulting polarization charge densities. The model for blood-brain partitioning was built on a data set of 103 compounds and yielded a correlation coefficient of r2 = 0.71 and an rms error of 0.40 log units. The model was validated by leave-one-out cross-validation tests, which resulted in q2 = 0.68 and a qms error of 0.42 for the logBB model.

The methodology for predicting the distribution of compounds between Blood and Brain, i.e. their brain/blood partition coefficients (logBB values), was stu-died using a non-linear regression analysis by Zhang and co-workers [129]. The equations were established on the basis of the different states (neutral, cationic and anionic) of the compounds distributing into the three main brain components (lipid, protein and water). The equations bear strong fitting and predictive power for the distribution of compounds (total set: n = 160, r = 0.906, s = 0.326; training set: n = 139, r = 0.908, s = 0.320; testing set: n = 21, r = 0.903, s = 0.297), and can describe the distribution of the different states of the compounds in the three components of the brain. The compounds in the dataset contained many different types, such as drug molecules, small structure-simple molecules, carboxylic acids and also alkaloids. Therefore the equations were very useful and instructional for the prediction of the compound distribution into the brain and blood.

On the basis of the thermodynamic condition of the passive transport across membranes, Keseru and Molnar demonstrated that brain penetration of molecules can be described as a function of solvation free energy [130]. The equation derived from 55 molecules showed a good predictive ability. Rapid calculation of Gsolv by the generalized Bohr surface area (GB/SA) continuum solvation model enabled us to develop a virtual screening tool on the high-throughput scale. Comparing the predictive power of Richter's GSOLV to methods in the literature, they concluded that their approach outperforms other log BB prediction tools, calculating log BB faster than 6 s/molecule.

## Metabolism

The metabolic fate of a compound depends on a large number of variables related to both the chemical itself (chemical structure, physicochemical properties, etc.) and the biological system (enzyme and its environment) [131] (Figure 2). A drug that is rapidly metabolized will require multiple daily dosing or continuous infusion to maintain an adequate therapeutic plasma level. Likewise, a highly stable drug, not readily metabolized and eliminated, could have a prolonged half-life, which might influence its safety.

A typical drug metabolism pathway is the oxidation of the drug (phase I oxidation), followed by conjugation of the oxidized moiety with highly polar molecules, such as glucose, sulfate, methionine, cysteine or glutathione (phase II conjugation).

The key enzymes for phase I oxidation are the isoforms of the cytochrome P450 (CYP) family. The major human CYP isoforms involved in drug metabolism are CYP1A2, CYP2A6, CYP2C9, CYP2C19, CYP2D6, CYP2E1 and CYP3A4, of which CYP3A4, CYP2C9, CYP2C19 and CYP2D6 are responsible for approximately 80% of the known oxidative drug metabolism reactions [132]. The key phase II enzymes include UDP-dependent glucuronosyl transferase (UGT), phenol sulfotranferase (PST), estrogen sulfotransferase (EST), and glutathione-S-transferase (GST). These enzymes also exist as multiple isoforms.

Several aspects of metabolic behavior such as biotransformation, binding to enzymes, and catalytic reaction, may be predicted *in silico* [133, 134] and several methods have been developed mostly based on the knowledge of the structure and mechanism of the enzymes (protein structure; 3D-structure and accessibility of the binding site; catalytic activity; mechanisms; specificity and regioselectivity), or based on the physicochemical properties of the compound (molecular sites sensitive to oxidation or

106. Bjorkman S. Prediction of the volume of distribution of a drug: which tissue-plasma partition coefficients are needed?. J Pharm Pharmacol 2002;54:1237-45.

107. Poulin P, Krishnan K. A biologically-based algorithm for predicting human tissue: blood partition coefficients of organic chemicals. Hum Exp Toxicol 1995;14:273-80.

108. Poulin P, Theil FP. Prediction of pharamcokinetics prior to *in vivo* studies. 1. Mechanism-based prediction of volume of distribution. J Pharm Sci 2002;91: 129-56.

109. Poulin P, Theil FP. Prediction of pharamcokinetics prior to *in vivo* studies. II. Generic physiologically based pharmacokinetic models of drug disposition. J Pharm Sci 2002;91:1358-70.

110. Lombardo F, Obach RS, Shalaeva MY, Gao F. Prediction of volume of distribution values in humans for neutral and basic drugs using physicochemical measurements and plasma protein binding data. J Med Chem 2002;45:2867-76.

111. Oie S, Tozer TN . Effect of altered plasma protein binding on apparent volume of distribution. J Pharm Sci 1979;68:1203-5.

112. Wichmann K, Diedenhofen M, Klamt A. Prediction of blood-brain partitioning and human serum albumin binding based on COSMO-RS sigma-moments. J Chem Inf Model 2007;47:228-33.

113. Gleeson MP, Waters NJ, Paine SW, Davis AM. *In silico* human and rat Vss quantitative structure-activity relationship models. J Med Chem 2006;49:1953-63.

114. Ghafourian T, Barzegar-Jalali M, Dastmalchi S, Khavari-Khorasani T, Hakimiha N, Nokhodchi A. QSPR models for the prediction of apparent volume of distribution. Int J Pharm 2006;319: 82-97.

115. Begley D J. The blood-brain barrier: Principles for targeting peptide and drugs to the central nervous system. J Pharm Pharmacol 1996;48:136-40.

116. Hansch C, Bjorkroth JP, Leo A. Hydrophobicity and centralnervous system agents: On the principle of minimal hydrophobicityin drug design. J Pharm Sci 1987;76:663-87.

117. van Bree JBMM, De Boer AG, Danhof M, Ginsel LA, Breimer DD. Characterization of an *in vitro* blood-brain barrier:Effects of molecular size and lipophilicity on cerebrovascular endothelial transport rates of drugs. J Pharmacol Exp Ther 1988;247:1233-9.

118. Kaliszan R, Markuszewski M. Brain/blood distribution described by a combination of partition coefficients and molecular mass. Int J Pharm 1996;45: 9-16.

119. Abraham MH, Chadha HS, Mitchel RC. Hydrogen bonding.33. Factors that influence the distribution of solutes between blood and brain. J Pharm Sci 1994;83:1257-68.

120. Abraham MH, Chadha HS, Mitchel RC. Hydrogen bonding. 36. Determination of blood brain distribution using octanol-water partition coefficients. Drug Des Discuss 1995;13:123-31.

conjugation, structure of the chemical, molecular surface properties, electronic structure, quantum mechanics properties, etc).

Due to the importance of cytochrome P450 in drug metabolism many models have been developed to predict both substrates and inhibitors. 3D models of cytochrome P450 using the x-ray structure of bacterial cytochromes P450 as templates, were developed [135]. By docking some substrates into the active site of their metabolizing enzymes it was found that substrates binding into the active site of isoform CYP2D6 are mainly favored by hydrogen bonding and electrostatic interactions between the substrate and two residues of the enzyme, in contrast, van der Waals attraction interactions mainly contribute to stabilize the complex involving both CYP1A2 and CYP3A4. The use of several molecular and intermolecular-interaction descriptors has also been proposed, for translating the qualitative structural information of substrate/cytochrome interaction models into semi-quantitative models of substrate specificity.

Pharmacophore models have been developed to assess the effects of 14 inhibitors on 7-benzyloxy-4-trifluoromethylcoumarin (BFC) metabolism by recombinant CYP3A4, CYP3A5 and CPY3A7, on directly comparing potential structural features and positioning differences for each enzyme. The CYP3A4 pharmacophore was characterized by the wide distance (14.3 A) between the furthest aromatic ring (hydrophobic) feature and the hydrogen-bond acceptor, whereas CYP3A5 and CYP3A7 contained the hydrogen-bond acceptor feature and a compact arrangement of three hydrophobic features. This could indicate that compounds that inhibit the metabolism of BFC by CYP3A4 are likely to have key hydrophobic interactions further away from the hydrogen-bond acceptor than in those molecules that inhibit CYP3A5 and CYP3A7 [136].

Inhibition of cytochrome P450 enzymes is unwanted because of the risk of severe side effects due to drug-drug interactions. Burton and co-workers explored the use of detailed biological data combined with a statistical learning method for predicting the CYP1A2 and CYP2D6 inhibition [137]. Data were extracted from the Aureus-Pharma highly structured databases which contain accurate measures and detailed experimental protocol concerning the inhibition of the two cytochromes. The methodology used was Recursive Partitioning. Models structuring was preceded by the evaluation of the chemical space covered by the datasets. The descriptors used are available in the MOE software suite. CYP2D6 datasets provided 11 models with an accuracy of over 80%, while CYP1A2 datasets had 5 high-accuracy models. Theses models can be useful to predict the ADME properties during the drug discovery process and are recomended for high-throughput screening.

Two *in silico* Gaussian kernel weighted k-nearest neighbor models based on extended connectivity fingerprints that classify CYP2D6 and CYP3A4 inhibition were developed by Jensen and co-workers [138]. Data used for modeling consisted of diverse sets of 1153 and 1382 drug candidates tested for CYP2D6 and CYP3A4 inhibition in human liver microsomes. For CYP2D6, 82% of the classified test set compounds was predicted for the correct class. For CYP3A4, 88% of the classified compounds were correctly classified. CYP2D6 and CYP3A4 inhibition were additionally classified for an external test set on 14 drugs, and multidimensional scaling plots showed that the drugs in the external test set were in the periphery of the training sets. Furthermore, fragment analysis were performed and structural fragments frequent in CYP2D6 and CYP3A4 inhibitors and non-inhibitors were found.

The software Catalyst has often been used for pharmacophore modeling drug metabolizing enzymes such as CYP2B6 [139], CYP2C9 [140], CYP2D6 [141], and CYP3A4 [142], and transporters such as P-glycoprotein [99, 143] and organic cation transporter [144]. CoMFA models have also been used for the modeling of inhibitors for enzymes like: CYP1A2 [145], CYP2A5 [146], CYP2A6 [146, 147], CYP2C9 [41, 148], CYP2D6 [149] and rat MAO A and b [150, 151].

In the case of phase II metabolism, only two X-ray crystal structures of UGT are recently available in the Protein Data Bank, UGT71G1 alone and in a complex with UDP-glucose [152].

Prior to the crystal structures, some *in silico* models were developed focused on the structures of UGT substrates, one of the first models of substrates for UGT were pharmacophore models of UGT1A1 [153] and UGT1A4 [154]. Common pharmacophore models demonstrated the importance of two hydrophobic domains separated from the glucuronidation site by 4 A and 7 A, respectively. The UGT1A1 and UGT1A4 models demonstrated useful predictive capability. However, difficulty in generating sufficient high-quality data, atypical glucuronidation kinetics, and possible multiple binding orientations of substrates within the UGT active site currently limit this approach [155].

Computer systems to predict xenobiotic metabolism are commercially available (Table 1). These softwares mainly predict phase I metabolism, although some of them also predict phase II reactions. COMPACT (Computer-Optimized Molecular Parametric Analysis of Chemical Toxicity) [156, 157] has modules that assess the ability of xenobiotics to form enzyme-substrate complexes and undergo metabolic activation by the CYP1A and CYP2E subfamilies of cytochromes P450. The COMPACT model uses sterical and electronic parameters. Improved criteria for CYP1A and CYP2E1 substrate specificity have recently been reported [158].

Camitro's metabolism models are based on a novel, combined empirical and quantum chemical approach to predict enzyme-substrate binding affinities, metabolic sites, and relative rates of metabolism at discrete sites within a molecule. The model is based on the evaluation of the energy necessary to attract a hydrogen atom from different groups, facilitating, for example, the calculation of aromatic oxidation or S-oxidation [39]. The main focus of the models are on the three major cytochrome P450 enzymes: CYP3A4, CYP2D6, and CYP2C9, which mediate over 90% of human drug metabolism, having other modules for absorption, based on a diffusion model, and blood-brain barrier partition based on the partition coefficient.

META [159, 160] is an expert system that can predict the sites of potential enzymatic attack and the nature of the chemicals formed by such metabolic

121. Norinder U, Sjoberg P, Ostenberg T. Theoretical calculation and prediction of brain-blood partitioning of organic solutes using Molsurf parametrization and PLS statistics. J Pharm Sci 1998;87:952-9.

122. Luco JM . Prediction of the brain-blood distribution of a large set of drugs from structurally derived descriptors using partial least-squares(PLS) modeling. J Chem Inf Comput Sci 1999;39:396-404.

123. Lombardo F, Blake JF, Curatolo WJ. Computation of brain-blood partitioning of organic solutes via free energy calculations. J Med Chem 1996;39:4750-5.

124. Ajay Bemis GW, Murcko MA . Designing libraries with CNS activity. J Med Chem 1999;42:4942-5.

125. Crivori P, Cruciani G, Carrupt P-A, Testa B. Predicting blood-brain barrier permeation from three-dimensional molecular structure. J Med Chem 2000; 43:2204-16.

126. Young RC, Mitchell RC, Brown TH, Ganellin CR, Griffiths R, Jones M, *et al*. Development of new physicochemical model for brain penetration and application to the design of centrally acting H2 receptor histamine antagonists. J Med Chem 1988;31:656-71.

127. Ter Laak AM, Tsai RS, Donne-Op den Kelder GM, Carrupt P-A, Testa B, Timmermann H . Lipophilicity and hydrogen-bonding capacity of H1-antihistaminic agents in relation to their central sedative side-effect. Eur J Pharm Sci 1994;2: 373-84.

128. Bendels S, Kansy M, Wagner B, Huwyler J. *In silico* prediction of brain and CSF permeation of small molecules using PLS regression models. Eur J Med Chem; In Press, Corrected Proof:.

129. Zhang H, Hu S, Zhang Y. Prediction of distribution of neutral, acidic and basic structurally diverse compounds between blood and brain by the nonlinear methodology. Med Chem 2008;4:170-89.

130. Keseru GM, Molnar L. High-throughput prediction of blood-brain partitioning: A thermodynamic approach. J Chem Inf Comput Sci 2001;41:120-8.

131. Kumar GN, Surapaneni S. Role of drug metabolism in drug discovery and development. Med Res Rev 2001;21: 397-411.

132. Williams JA, Hyland R, Jones BC, Smith DA, Hurst S, Goosen TC, Peterkin V, Koup JR, Ball SE. Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. Drug Metab Dispos 2004;32: 1201-8.

133. Ekins S, Waler CL, Swaan PW, Cruciani G, Wrighton SA, Wikel JH . Progress in predicting human ADME parameters *in silico*. J Pharmacol Toxicol Methods 2000;44:251-72.

134. Ekins S, Wrighton SA. Application of *in silico* approaches to predicting drug-drug interactions. J Pharmacol Toxicol Methods 2001;45:65-9.

135. De Rienzo F, Fanelli F, Menziani MC, De Benedetti PG. Theoretical investigation of substrate specificity for cytochromes P450 IA2, P450 IID6 and P450 IIIA4. J Comput Aided Mol Des 2000;14: 93-116.

transformations. The program uses dictionaries of biotransformation operators created by experts in the field of xenobiotic metabolism to represent known metabolic pathways [159]. Activation of a biotransformation operator within the program is based on the recognition of key functional groups within the complete chemical structure [160]. The program evaluates stable metabolites and transform them further build up a complete metabolic tree. Processing continues in this way until the program is able to generate water soluble metabolites that are to be excreted.

MetabolExpert predicts the metabolism of exogenous compounds in plants and mammals. The system is composed of biotransformation databases, a knowledge-base and prediction tools [161]. The transformation knowledge-base is based on 'if/then' type rules derived from the literature and the users can modify or delete a given rule. Each biotransformation rule includes four components: the substructure changed during the transformation, the new substructure formed during the transformation, a list of substructures at least one of which must be present in the molecule for the biotransformation to occur and a list of substructures whose presence prevents the transformation from occurring. There are two types of predictions in MetabolExpert. In the first, the system tries to match basic biotransformations to the compound structure. The second type of analysis is an extended prediction model in which metabolites generated from basic transformations are compared to transformations in 'learned' trees.

METEOR [162], developed by LHASA is a computer system which uses a knowledge-base of structure-metabolism rules (biotransformations) to predict the metabolic fate of a query chemical structure [161]. The reasoning engine in METEOR takes into account the knowledge of chemical reaction mechanisms, lipophilicity, competition between possible reactions, etc. METEOR's biotransformation rules are generic reaction descriptors rather than simple entries in a reaction database. The users can build their own biotransformation and rules. The system uses a rich internal structure representation language [163], so the expression of specific functional group transformations can be made context-sensitive. The only information needed by the program to make the prediction is the molecular structure of the compounds. The system can compare potentially competing biotransformations [163] and the user can choose to analyze queries at a number of available search levels. METEOR and MetabolExpert are considered easy to use, give rapid answers and are linked to a toxicity prediction system [164]. They enable the interaction of the user in the generation of the metabolic tree, namely to distinguish between phase I and phase II reactions. These programs and their combinations usually predict many more metabolites than those observed experimentally; nevertheless some metabolites observed *in vivo* are not correctly predicted.

## Excretion

Drugs are eliminated from the body by metabolism and excretion. The most common routes of elimination are renal and / or bile excretion. The kidney is the major contributor to drug excretion and also to the excretion of certain metabolites. Large molecules and many drugs conjugated with glucuronic acid are generally excreted in the bile. Drugs eliminated in the bile are available for re-absorption in the gastrointestinal tract. This re-absorption after their 'elimination' through the bile results in the 'recycling' of drugs and prolongs the time required for the drug to be irreversibly eliminated.

The kidney has developed high capacity transport systems to rapidly eliminate the large amounts of foreign compounds delivered to it. Even drugs that are 'eliminated' only by hepatic metabolism often depend on renal excretion for the ultimate elimination of the xenobiotic from the body. Filtration through the glomerular membranes is the main mechanism for drug and metabolite excretion into the urine. However, some drugs are excreted by active transport mechanisms serving as substrates for tubular or biliary epithelial transporters. In recent years, much study has been focused on understanding the cellular and molecular basis of drug transport systems responsible for this elimination [165, 166]. Active transport systems such as organic anion transporting polypeptides, peptide transporters, organic anion transporters and organic cation transporters can also carry compounds and/or their metabolites across the basolateral surface of renal or hepatic membranes, resulting in drug elimination [165]. Organic anion transporters can also bring drugs from the blood into the renal tubular cells and organic cation transporters transfer the drug from inside the cells into the tubular lumen [166]. Other families of transporters, such as: P-gp, BCRP and MDR, mediate the efflux across apical surfaces resulting in secretion and elimination.

There has been very little work on the *in silico* modeling or prediction of excretion [164]. Passive excretion can theoretically be predicted using some of the approaches for the prediction of tissue distribution, since it is determined by similar physicochemical and physiological properties (blood flow, protein binding, lipophilicity, pKa ), possibly with different limits, e.g. glomerular filtration and molecular weight. However, in practice metabolic stability *in vitro* / *in vivo* and initial animal pharmacokinetic studies would give the first indications on the potential significance of the renal excretion route [164]. *In silico* modeling of the P-gp substrates and inhibitor can be used to predict excretion because it acts as an effluent pump from inside the renal cell into the collecting duct [166].

## Toxicity

Drug toxicity is a property of paramount importance. A desirable drug would have a high therapeutic index; being the plasma level required for exerting a toxic effect, significantly higher than that required for therapeutic efficacy. Human toxicity continues to occur in clinical trials of drug candidates that are apparently found to be safe during preclinical trials.

There are also indications that some substructures found in drugs can form reactive metabolites involved in toxicity. These substructures include arylacetic and arylpropionic acids, aryl hydroxamic acids, oximes, anilines, anilides, hydrazines, hydrazides, hydantoins, quinones, quinone methides, nitroaromatics, hetero-

136. Ekins S, Stresser DM, Williams JA. *In vitro* and pharmacophore insights into CYP3A enzymes. Trends Pharmacol Sci 2003;24:161-6.

137. Burton J, Ijjaali I, Barberan O, Petitet F, Vercauteren DP, Michel A. Recursive partitioning for the prediction of cytochromes P450 2D6 and 1A2 inhibition: importance of the quality of the dataset. J Med Chem 2006;49:6231-40.

138. Jensen BF, Vind C, Padkjaer SB, Brockhoff PB, Refsgaard HH. *In silico* prediction of cytochrome P450 2D6 and 3A4 inhibition using Gaussian kernel weighted k-nearest neighbor and extended connectivity fingerprints, including structural fragment analysis of inhibitors versus noninhibitors. J Med Chem 2007; 50:501-11.

139. Ekins S, Bravi G, Ring BJ, Gillespie TA, Gillespie JS, Vandenbranden M, *et al*. Three-dimensional quantitative structure activity relationship analyses of substrates for CYP2B6. J Pharmacol Exp Ther 1999; 288:21-9.

140. Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH, Wrighton SA. Three- and four-dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2C9 inhibitors. Drug Metab Dispos 2000;28:994-1002.

141. Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH, Wrighton SA. Three and four dimensional-quantitative structure activity relationship (3D/4D-QSAR) analyses of CYP2D6 inhibitors. Pharmacogenetics 1999;9:477-89.

142. Ekins S, Bravi G, Wikel JH, Wrighton SA. Three-dimensional-quantitative structure activity relationship analysis of cytochrome P-450 3A4 substrates. J Pharmacol Exp Ther 1999;291:424-33.

143. Ekins S, Kim RB, Leake BF, Dantzig AH, Schuetz EG, Lan LB, *et al* . Application of three-dimensional quantitative structure-activity relationships of P-glycoprotein inhibitors and substrates. Mol Pharmacol 2002;61:974-81.

144. Bednarczyk D, Ekins S, Wikel JH, Wright SH. Influence of molecular structure on substrate binding to the human organic cation transporter, hOCT1. Mol Pharmacol 2003;63:489-98.

145. Fuhr U, Strobl G, Manaut F, Anders EM, Sorgel F, Lopez-de-Brinas E, Chu DT, *et al*. Abstract Quinolone antibacterial agents: relationship between structure and *in vitro* inhibition of the human cytochrome P450 isoform CYP1A2. Mol Pharmacol 1993;43:191-9.

146. Poso A, Gynther J, Juvonen R. A comparative molecular field analysis of cytochrome P450 2A5 and 2A6 inhibitors. J Comput Aided Mol Des 2001;15:195-202.

147. Asikainen A, Tarhanen J, Poso A, Pasanen M, Alhava E, Juvonen RO. Predictive value of comparative molecular field analysis modelling of naphthalene inhibition of human CYP2A6 and mouse CYP2A5 enzymes. Toxicol *In Vitro* 2003; 17:449-55.

148. Rao S, Aoyama R, Schrag M, Trager WF, Rettie A, Jones JP. A refined 3-dimensional QSAR of cytochrome P450 2C9: computational predictions of drug interactions. J Med Chem 2000;43: 2789-99.

aromatics, halogenated hydrocarbon, some halogenated aromatics, which are chemical groups that can be oxidized to acroleins, and medium-chain fatty acids [167].

Toxicity prediction is a wearisome task; it could be species-specific, organ-specific (Figure 2), and could involve multiple host factors and chronic dosing regimes, all of which cannot be adequately modeled experimentally. Hepatotoxicity is a major manifestation of drug toxicity, and it is known that toxicity can be influenced by drug metabolism, therefore screening for toxicity using intact hepatocytes is an approach being adopted in many drug discovery and development laboratories.

The human Ether-a-go-go Related Gene (hERG) potassium channel blockage has become a growing concern for both regulatory agencies and pharmaceutical industries who invest substantial effort in the assessment of the cardiac toxicity of drugs. I*n silico* tools to filter out potential hERG channel inhibitors at early stages of the drug discovery process have considerable potential for saving time and money, since patch-clamp measurements are very expensive and no crystal structures of the hERG-encoded channel are available.

A predictive QSAR model for hERG blockade was developed by Kramer and co-workers [168]. This model differentiates the specific and nonspecific binding. Specific binders are identified by preliminary pharmacophore scanning. In addition to descriptor-based models for the compounds selected as hitting one of two different pharmacophores, they also use a model for nonspecific binding that reproduces blocking properties of molecules that do not fit into either of the two pharmacophores. PLS and SVR (Support Vector Regression) models based on interpretable quantum mechanics derived descriptors on a literature dataset of 113 molecules reach overall R(2) values of between 0.60 and 0.70 for independent validation sets and R(2) values of between 0.39 and 0.76 after partitioning according to the pharmacophore search for the test sets. These findings suggest that hERG blockade may occur through different types of binding, so that several different models may be necessary to assess hERG toxicity [168].

Li and co-workers developed binary classification models based on a large and diverse library of 495 compounds. The models combine pharmacophore-based GRIND descriptors with a support vector machine (SVM) classifier in order to discriminate between hERG blockers and non-blockers [169]. The models were applied at different thresholds from 1 to 40 microns. The model at a 40 micron threshold showed the best performance and was validated internally (Matthew's coefficient correlation of 0.40 and F-measure of 0.57 for blockers and 0.81 for non-blockers, using a leave-one-out cross-validation). On an external set of 66 compounds, 72% was correctly predicted (F-measure of 0.86 and 0.34 for blockers and non-blockers, respectively). The model was also tested on a large set of hERG bioassay data recently made publicly available on PubChem (http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=376) to achieve an accuracy of about 73% (F-measure of 0.30 and 0.83 for blockers and non-blockers, respectively). Even if there is still some limitation in the assessment of the hERG blockers, the performance of this model shows an improvement of between 10% and 20% in the prediction of blockers compared to other methods, which can be useful in screening potential hERG channel inhibitors.

Some of available expert systems for toxicity prediction are VirtualToxLab, OncoLogic, MultiCASE, TOPKAT, DEREK for Windows and HazardExpert (Table 1).

VirtualToxLab includes a 6D-QSAR concept and validated a series of 10 virtual test kits based on the aryl hydrocarbon, estrogen alpha/beta, androgen, thyroid alpha/beta, glucocorticoid and peroxisome proliferator-activated receptor gamma as well as on the enzymes CYP3A4 and CYP2A13, respectively [170, 171]. The test kits are based on the three-dimensional structure of their target protein (*i.e*. ER (alpha/beta), AR, TR(alpha/beta), CYP450) or a surrogate thereof (AhR) and were trained using a representative selection of 362 substances [170]. The test kits were trained using a representative selection of 628 substances and validated with 194 compounds different therefrom. The surrogates have been tested against a total of 798 compounds and are able to predict a binding affinity close to experimental uncertainty, with only six of the 188 test compounds calculated having more than a factor of 10 off the experimental binding affinity and a maximal individual deviation that does not exceed a factor of 15. These results suggest that this approach is suited for the *in silico* identification of adverse effects triggered by drugs and environmental chemicals. An Internet Portal to the VirtualToxLab was also developed to provide easy access to this technology [171]. Upon uploading the 3D coordinates of one or more compounds of interest, the compounds will be automatically processed and tested against the selected virtual test kits by means of consensus scoring. The user can interactively monitor the simulation process and have access to all results.

The OncoLogic Cancer Expert System was developed under a cooperative agreement between EPA's Office of Pollution Prevention and Toxics (OPPT) and LogiChem Inc. OncoLogic can analyze a chemical structure to determine the likelihood that it may cause cancer; this is done by applying rules of structure activity relationship (SAR) analysis and incorporating certain knowledge of chemicals that could cause cancer in animals and humans. This program is being released by EPA at no cost and runs on a Windows PC.

The MultiCASE (multiple computer automated structure evaluation) [172-174] approach evaluates a data set by trying to identify the structural features responsible for the activity (biophores). It then creates organized dictionaries of these biophores and develops QSAR correlations that can be used to predict the activity of unknown molecules. A new molecule will be evaluated against the dictionary and the appropriate QSARs it has created and, based on the results, it gives a prediction for the projected activity of the molecule in the corresponding test.

The TOPKAT (Toxicity Prediction by Komputer Assisted Technology) [175, 176] system predicts toxicity endpoints based on chemical structure, including rodent carcinogenicity, Ames mutagenicity, rat

149. Haji-Momenian S, Rieger JM, Mac donald TL, Brown ML. Comparative molecular field analysis and QSAR on substrates binding to cytochrome p450 2D6. Bioorg Med Chem 2003;11:5545-54.

150. Medvedev AE, Veselovsky AV, Shvedov VI, Tikhonova OV, Moskvitina TA, Fedotova OA, *et al*. Inhibition of monoamine oxidase by pirlindole analogues: 3D-QSAR and CoMFA analysis. J Chem Inf Comput Sci 1998;38:1137-44.

151. Moron JA, Campillo M, Perez V, Unzeta M, Pardo L. Molecular determinants of MAO selectivity in a series of indolylmethylamine derivatives: biological activities, 3D-QSAR/CoMFA analysis, and computational simulation of ligand recognition. J Med Chem 2000;43:1684-91.

152. Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X. Crystal Structures of a Multifunctional Triterpene/Flavonoid Glycosyltransferase from Medicago truncatula. Plant Cell 2005;17:3141-54.

153. Sorich MJ, Smith PA, McKinnon RA, Miners JO. Pharmacophore and quantitative structure activity relationship modelling of UDP-glucuronosyltransferase 1A1 (UGT1A1) substrates. Pharmacogenetics 2002;12:635-45.

154. Smith PA, Sorich MJ, McKinnon RA, Miners JO. Pharmacophore and quantitative structure activity relationship (QSAR) modelling: complemetary approaches to the rationalization and prediction of UDP-gluconosyltransferase 1A4 (UGT1A4) substrate selectivity. J Med Chem 2002; 46:1617-26.

155. Miners JO, Smith PA, Sorich MJ, Mc Kinnon RA, Mackenzie PI. Predicting human drug glucuronidation parameters: Application of *in vitro* and *in silico* modelling approaches. Annu Rev Pharmacol Toxicol 2004;44:1-25.

156. Lewis CVF, Ioannides C, Parke DV. An improved and updated version of the compact procedure for the evaluation of P450-mediated activation. Drug Metab Rev 1998;30:709-37.

157. Lewis DVF. COMPACT: a structural approach to the modelling of cytochromes P450 and their interactions with xenobiotics. J Chem Technol Biotechnol 2001;76:237-44.

158. Lewis DVF, Ioannides C, Parke DV. Further evaluation of COMPACT, the molecular orbital approach for the prospective safety evaluation of chemicals. Mutat Res 1998; 412:41-54.

159. Klopman G, Dimayuga M, Talafous J. META 1. A program for the evaluation of me-tabolic transformation of chemicals. 1994; 34:1320-5.

160. Talafous J, Sayre LM, Mieyal JJ, Klopman G . META 2. A dictionary model of mammalian xenobiotic metabolism. J Chem Inf Comput Sci 1994;34:1326-33.

161. Langowski J, Long A. Computer systems for the prediction of xenobiotic metabolism. Adv Drug Deliv Rev 2002; 54407-415.

162. Greene N, Judson PN, Langowski JJ, Marchant CA. Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. SAR QSAR Environ Res 1999;10:299-314.

oral LD50, rat chronic lowest-observable adverse effect level, developmental toxicity potential and skin sensitization. These predictions are achieved through sets of biostatistic expressions derived from well-characterized databases. The QSAR model uses two-dimensional molecular, electronic and spatial descriptors. Toxicity models can be used with large 100,000-plus data sets for virtual high-throughput screening and library design, helping to identify toxicity problems at the identification and optimization stages of the drug development process.

Deductive Estimation of Risk from Existing Knowledge for Windows (DEREK for Windows) [162, 177, 178] provides a qualitative assessment of toxicity potential using structure-based alerts that define toxophores. The program applies structure-activity relationships ((Q)SARs) and other expert knowledge rules, with an emphasis on the understanding of mechanisms of toxicity and metabolism, to derive a reasoned conclusion about the potential toxicity of the query chemical. The system is able to perceive chemical sub-structures within molecules and relate these to a rule-base linking the sub-structures with likely types of toxicity.

HazardExpert is a toxicity prediction system based on the structure of the compounds and predicts different effects such as: carcinogenic, mutagenic, teratogenic, and neurotoxic. HazardExpert gives toxicokinetic and toxicodynamic investigation of the whole molecule by calculating its molecular weight, pKa and logP values. The bioavailability of the compounds can also be considered by a simple but powerful model. In some cases, the drug's metabolites are responsible for the toxic effect. To predict the toxicity of both the parent compound and the metabolites, some toxicity prediction programs can be linked with metabolism by a metabolism prediction program. For example, DEREK can be linked with METEOR, and Hazard Expert has been used together with the COMPACT system for evaluating human carcinogenicity data [179].

## Conclusions

It is clear from the literature that there is neither a general methodology nor a trend for the computer prediction of ADME-Tox properties. A number of programs are available and can be used for the estimation of several physicochemical properties and many of them have been around for many years. Methods and algorithms based on very specific molecular descriptors are now being developed. In general the main limitation for the development of new methods and for the usage of the more general available programs is still the lack of enough chemical, physicochemical and biochemical data for a thorough statistical assessment of the quality of the models, making the examples of a successful application of the ADME-Tox models, exceptional and very elaborated.

163. Tonnelier CAG, Fox J, Judson PN, Krause PJ, Pappas M, Patel N. Representation of chemical structures in knowledge-based systems: the StAR system. J Chem Inf Comput Sci 1997; 37:117-23.

164. Boobis A, Gundert-Remy U, Kremers P, Macheras P, Pelkonen O. In silico prediction of ADME and pharmacokinetics. Report of an expert meeting organised by COST B15. Eur J Pharm Sci 2002;17:183-93.

165. Lee W, Kim RB. Transporters and renal drug elimination. Ann Rev Pharmacol Toxicol 2004;44:137-66.

166. Launay-Vacher V, Izzedine H, Karie S, Hulot JS, Baumelou A, Deray G. Renal tubular drug transporters. Nephron Physiol 2006; 103:97-106.

167. Nassar AEF, Kamel AM, Clarimont C. Improving the decision-making process in structural modification of drug candidates: reducing toxicity. Drug Discov Today 2004; 9:1055-64.

168. Kramer C, Beck B, Kriegl JM, Clark T. A Composite Model for hERG Blockade. Chem Med Chem 2008;3:254-65.

169. Li Q, Jorgensen FS, Oprea T, Brunak S, Taboureau O. hERG classification model based on a combination of support vector machine method and GRIND descriptors. Mol Pharm 2008;5:117-27.

170. Vedani A, Dobler M, Lill MA. The challenge of predicting drug toxicity *in silico*. Basic Clin Pharmacol Toxicol 2006;99:195-208.

171. Vedani A, Dobler M, Spreafico M, Peristera O, Smiesko M. VirtualToxLab - *in silico* prediction of the toxic potential of drugs and environmental chemicals: evaluation status and internet access protocol. ALTEX 2007; 24:153-61.

172. Klopman G. MULTI-CASE. 1. A hierarchical computer automated structure evaluation program. Quant Struct Act Relatsh 1992; 11:176-84.

173. Rosenkranz HS, Cunningham AR, Zhang YP, Claycamp HG, Macina OT, Sussman MB, *et al*. Development, characterization and application ofpredictive toxicology models. SAR QSAR Environ Res 1999;10:277-98.

174. Klopman G, Chakravarti SK, Harris N, Ivanov J, Saiakhov RD. *In silico* screening of high production volume chemicals for mutagenicity using the MCASE QSAR expert system. SAR QSAR Environ Res 2003;14: 165-80.

175. Prival MJ. Evaluation of the TOPKAT system for predicting the carcinogenicity of chemicals. Environ Mol Mutagen 2001; 37:55-69.

176. Enslein K, Gombar VK, Blake BW. Internation Commission for Protection Against Environmental Mutagens and Carcinogens. Use of SAR in computer-assisted prediction of carcinogenicity and mutagenicity of chemicals by the TOPKAT program. Mutat Res 1994;305:47-61.

177. Sanderson DM, Earnshaw CG. Computer prediction of possible toxic action from chemical structure the DEREK system. Hum Exp Toxicol 1991;10:261-73.

178. Ridings JE, Barratt MD, Cary R, Earnshaw CG, Eggington CE, Ellis M, *et al*. Computer prediction of possible toxic action from chemical structure: an update of the DEREK system. Toxicology 1996;8:267-79.

179. Lewis DF, Ioannides C, Parke DV. A combined COMPACT and HazardExpert study of 40 chemicals for which information on mutagenicity and carcinogenicity is known, including the results of human epidemiological studies. I 1998;17:577-86.

*Received in October, 2007. Accepted for publication in May, 2008.*