QSAR and QM/MM Approaches Applied to Drug Metabolism Prediction

R.C. Braga and C.H. Andrade*

Laboratory of Molecular Modeling and Design (LabMol), Faculty of Pharmacy, Federal University of Goiás, Avenida Universitária com 1a. Avenida, PO Box 131, Goiânia, GO, 74605-220, Brazil

Abstract: In modern drug discovery process, ADME/Tox properties should be determined as early as possible in the test cascade to allow a timely assessment of their property profiles. To help medicinal chemists in designing new compounds with improved pharmacokinetics, the knowledge of the soft spot position or the site of metabolism (SOM) is needed. In recent years, large number of *in silico* approaches for metabolism prediction have been developed and reported. Among these methods, QSAR models and combined quantum mechanics/molecular mechanics (QM/MM) methods for predicting drug metabolism have undergone significant advances. This review provides a perspective of the utility of QSAR and QM/MM approaches on drug metabolism prediction, highlighting the present challenges, limitations, and future perspectives in medicinal chemistry.

Keywords: QSAR, QM/MM, drug metabolism prediction, cytochrome P450, docking.

INTRODUCTION

In modern drug discovery process, ADME/Tox (absorption, distribution, metabolism, excretion, along with toxicity) properties should be determined as early as possible in the test cascade to allow a timely assessment of their property profiles [1]. Among ADME/Tox properties, drug metabolism is a key determinant of several important drug processes *in vivo*, such as metabolic stability, drug–drug interactions and drug toxicity [2].

Metabolism is the biochemical transformation of a drug or xenobiotic, and it is traditionally divided into Phase I and Phase II processes. The first one involves the modification of a functional group by oxidation, reduction, or hydrolysis. Phase II of metabolism involves conjugation reactions, where a highly hydrophilic moiety such as sulfate or glucuronic acid is attached to make the drug more water-soluble and to prepare for excretion through urine or bile [3]. The most important group of Phase I enzymes is the cytochrome P450 (CYP) superfamily. Of the 57 known human isoforms of CYPs, just six seem to play a significant role in drug metabolism. In particular, the CYP1A2, 2C9, 2C19, 2D6, 2E1 and 3A4 isoenzymes account for the oxidative metabolism of > 90% of the drugs in the market [4]. Key Phase II enzymes include, for instance, uridine diphosphate-dependent glucuronosyl transferase (UGT), sulfotransferases and glutathione-*S*-transferase. Glucuronidation of small lipophilic molecules by UGTs is probably the most important Phase II process for the clearance of drugs [5].

CYPs and UGTs, which both exist as enzyme "superfamilies", are together responsible for the metabolism of most hepatically cleared drugs. There is currently intense interest in the development of techniques that permit identification of the CYP and UGT isoform(s) involved in the metabolism of a newly discovered drug, and hence prediction of factors likely to alter elimination *in vivo* [5, 6].

Therefore, it is crucial to have reliable information on how a chemical entity behaves in the presence of metabolizing enzymes. A key task is the identification of metabolites and, ideally, the modification of the compound to improve the pharmacokinetic or pharmacological profiles by changing the metabolic susceptibility. Consequently, being able to predict the likely site of metabolism (SOM) in any compound, synthesized or virtual, would be extremely useful.

In recent years, a large number of *in silico* or computational approaches for metabolism prediction have been developed and reported; these are mainly divided into structure-based and ligand-based approaches, also known as knowledge-based approach [7-9]. The structure-based approaches rely upon the structural information extracted from the X-ray crystallographic and/or homology drug metabolizing proteins such as CYPs [9, 10]. These approaches include 3D molecular modeling between the ligand and CYPs [6, 11, 12], quantum mechanical (QM) methods [13-16] and pharmacophore modeling [12, 17-20]. Knowledge or ligand-based approaches rely on the assumption that the metabolic fate of a compound is exclusively a consequence of its chemical structure and characteristics. All ligand-based models provide indirect information about a protein' active site on the basis of the shape, electronic properties and conformations of substrates, inhibitors or metabolic products. These models are often dependent on the availability of experimental data for a sufficiently large number of substrates. These approaches include various descriptor-based methods (e.g., classificators, structural similarities, quantitative structure–activity/property relationships or QSAR/QSPR, three-

*Address correspondence to this author at the LabMol, Faculty of Pharmacy, Federal University of Goiás, Avenida Universitária com 1a. Avenida, s/n, PO Box 131, Goiânia, GO, 74605-220, Brazil;
Tel: + 55 62 3209-6451; Fax: +55 62 3209 6037;
E-mail: carolina@farmacia.ufg.br

1875-5607/12 $58.00+.00

dimensional QSAR (3D-QSAR), quantum mechanical (QM) methods and pharmacophore. QM methods use calculations on various levels of theory to describe the electronic structure of ligands and to calculate the energies of various species along a given reaction.

Among these methods, QSAR models and combined quantum mechanics/molecular mechanics (QM/MM) methods for predicting drug metabolism have undergone significant advances recently. Fig. (1) presents a basic search on ISI Web of Knowledge for the combined queries "QM/MM" and "P450" or "QSAR" and "P450", showing a marked increase in publications over the past few years. This review provides a perspective of the utility of QSAR and QM/MM approaches on drug metabolism prediction, especially by means of P450 metabolism, highlighting the present challenges, limitations, and future perspectives in medicinal chemistry.

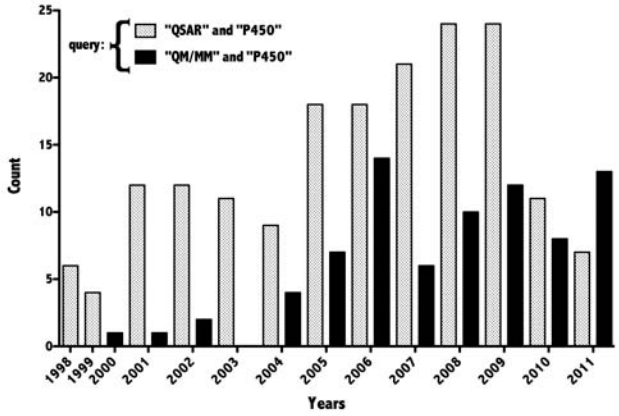

Fig. (1). Number of publications in the last 14 years matching the queries "QSAR" and "P450" (dotted bars), and "QM/MM" and "P450" (black bars). Source from ISI Web of Knowledge.

QSAR APPROACHES

In the drug design process, the methodology currently known as Quantitative Structure-Activity Relationship (QSAR) was definitively launched in the early 1960's with the innovative works of Hansch and Fujita [21] and Free and Wilson [22]. The underlying theory of QSAR methodology is that the differences observed in the biological activity are related to molecular structure [23, 24]. Therefore, biological activity of congeneric molecular structures can be mathematically expressed as a function of specific structural molecular features (descriptors) by using regression techniques to estimate the relative importance of those features contributing to the biological effect.

The classical QSAR methods [21, 25] use as molecular descriptors global molecular properties of ligands (e.g. pKa, logP, etc.) and/or those correlated with the 2D structural patterns (e.g. connectivity, 2D pharmacophore, etc.).

When the study of the 3D molecular structure became practical routine with the parallel development of several computational techniques in the 1980s, the new era of the drug design process, named Computer-Assisted Drug Design (CADD) came into being and QSAR methodology has come in a broad subfield of CADD [24, 26]. Since then, several QSAR methodologies have been proposed. The introduction of CoMFA (Comparative Molecular Fields Analysis) [27] in 1988 represents a milestone in QSAR as, for the first time, such structure-activity relationships were based on the 3D structure of the ligands (3D-QSAR). In CoMFA the ligands' interaction with chemical probes is mapped onto a surface or grid surrounding a series of compounds (superimposed in 3D space). This surface or grid represents a surrogate of the binding site of the true biological receptor.

The QSAR formalisms can be characterized by having particular approaches for calculating and selecting the molecular descriptors, and specific statistical algorithms for constructing the resulting models. Based on their dimensionality, QSAR approaches can be classified as follows: classical (zero-dimensional or 0D), one-dimensional (1D), two-dimensional (2D) [21, 22], three-dimensional (3D) [27], and four-dimensional (4D) [28, 29] QSAR approaches. The descriptors can be molecular features, such as atom and molecular counts, molecular weight, sum of atomic properties (0D-QSAR); fragment counts (1D-QSAR); topological descriptors (2D-QSAR); geometrical, atomic coordinates, or energy grid descriptors (3D-QSAR); and the combination of atomic coordinates and sampling of conformations (4D-QSAR) [26]. These methodologies comprise the receptor-independent (RI) analyses [26, 30]. This QSAR group is characterized by the construction of models in the absence of a well-defined structure for the molecular target. The other approach is the receptor-dependent (RD) analysis, in which models are derived from the 3D structure of the multiple ligand-receptor complex conformations [23, 30, 31]. This approach provides an explicit simulation of the induced-fit process, using the structure of the ligand-receptor complex, where both ligand and receptor are allowed to be completely flexible by the use of molecular dynamics (MD) simulation. RD-QSAR is used to gather binding interaction energies, as descriptors, from the interaction between the analog molecules and the receptor [31]. Due to the intrinsic dependence of atomic coordinates of both receptor and ligands, RD-QSAR includes multidimensional methods (*n*D-QSAR), such as 4D [32], 5D [33], and 6D [34] QSAR, among others.

QSAR MODELS FOR DRUG METABOLISM PREDICTION

QSAR models for predicting drug metabolism have undergone significant advances recently. Numerous descriptors, ranging from fragment-encoding fingerprints to physicochemical descriptors and descriptors computed from the spatial arrangement of pharmacophoric interaction points are available [35].

As shown in Fig. (2), QSAR models for predicting drug metabolism can be divided into four main steps, i.e., (*i*) determination or collection of biological property of interest (metabolism parameters), (*ii*) molecular descriptor generation and variable selection to extract desirable independent variables, (*iii*) model generation and validation with training and test sets using linear or non-linear

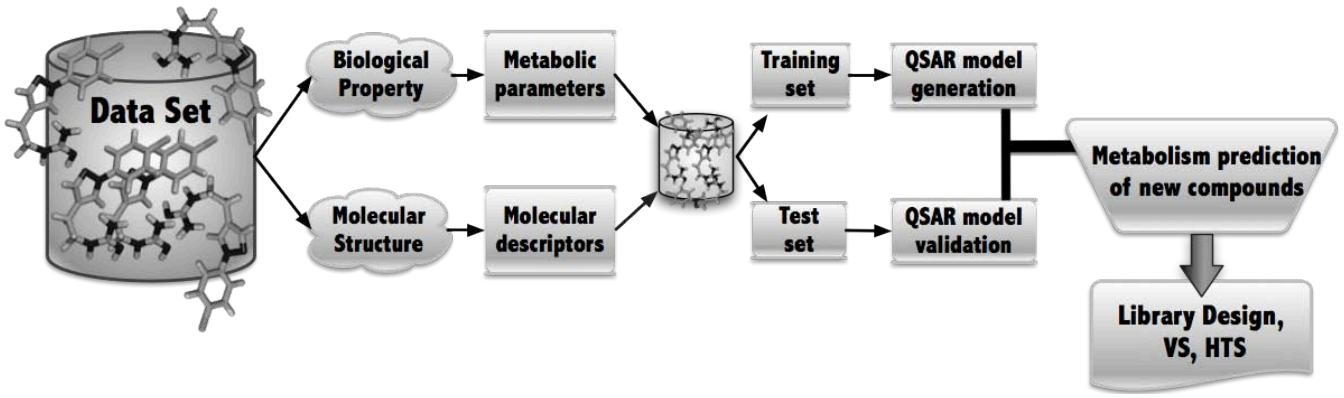

Fig. (2). Schematic workflow for construction of QSAR models for drug metabolism prediction.

statistical methods and, (*iv*) prediction of the metabolism of new compounds using an external validation set [36].

The generated QSAR models can be employed as metabolism filters in the process of chemical library design, virtual screening (VS) and high-throughput screening (HTS), therefore, integrating the study of pharmacodynamic and pharmacokinetic properties in the identification of new lead candidates.

Biological Property

It is essential to choose the type of biological property available to construct the QSAR models for drug metabolism prediction [36, 37]. Unlike the prediction of absorption and toxicity, for which the endpoint is relatively easy to select, such as the fraction absorbed from the intestine (or membrane permeability coefficient) and LD_{50} (medial lethal dose), there is no straightforward assay process for metabolism [18]. Basically, for the categorical model, outputs may be inhibitors and/or inducers, and substrates and/or non-substrates of CYP450; while the IC_{50} (50% inhibiting concentration), V_{max} (maximum rate of metabolism), hepatic metabolic clearance (CL_h) and *in vitro* intrinsic clearance ($CL_{int,\ in\ vitro}$) can be used as properties for a quantitative metabolism model [37]. Therefore, suitable metabolism parameters should be selected first, and should be care of accuracy/consistency of the entire data set.

Diversity of the Data Set

The rational division of the entire data set into training and test sets is one of the most important steps governing the predictability of a QSAR model [38]. Fundamentally, the whole data set can be divided into training and test sets in a random manner. Cluster-based methods, such as the *K*-means clustering algorithm, have also been used to create diverse training sets and representative test sets. Concomitantly, evaluation of the data set diversity by dissimilarity-based methods (DI), and the representativeness between training and test sets by representativeness index (RI), should be performed. However, until now, most of the QSAR studies involving drug metabolism prediction have not used the DI and RI values to evaluate the diversity of their modeling data sets.

Descriptors for the Prediction of Metabolism

In the process of QSAR model construction, various rationally designed molecular descriptors are needed to examine molecular structures. Different descriptors emphasize different chemical properties implicit in the molecular structure, usually divided into two-dimensional (2D), which encodes the topology of a molecule; and three-dimensional (3D), based on the 3D structures of a molecule [39]. The 2D descriptors are independent of the 3D orientation of drugs, including constitutional, electronic, quantum chemical, topological, geometrical descriptors, fragment-based descriptors and fingerprints. The 3D descriptors use distance information derived from spatial arrangement of atoms or atom groups, *i.e.* molecular conformation. The conformation is then refined by minimizing the energy [61] and, subsequently, the alignment of the conformers uniformly in space is performed. Finally, the space with immersed conformer is probed computationally for various descriptors. CoMFA with electrostatic and steric energy fields [40] and comparative molecular similarity indices (CoMSIA) [41] with steric, electrostatic, hydrophobic and H-bond donor or acceptor properties are commonly used for alignment-dependent 3D descriptors. Some 3D descriptors are derived independently of the molecular alignment, such as VolSurf [42] approach, and grid-independent descriptors (GRIND) [43]. The 4D descriptors are the grid cell occupancy descriptors (GCODs), which are generated for a number of different interaction pharmacophore elements (IPEs). These IPEs (i.e., atom types), defined as “any type” (A or Any), “nonpolar” (NP), “polar-positive charge” (P+), “polar-negative charge” (P-), “hydrogen bond acceptor” (HA), “hydrogen bond donor” (HB), and “aromatic” (Ar), correspond to the interactions that may occur in the active site, and are related to the pharmacophore groups [23, 28].

Descriptor Selection and QSAR Model Development

One of the major challenges in a QSAR is the selection of relevant molecular descriptors from large number of descriptors. Therefore, selection of proper and interpretable descriptors to establish QSAR models is a very important step to reduce over-fitting, speed up training, improve the overall model predictability, and to interpret the QSAR

model. At the same time, this is also a challenging and difficult step.

Variable selection can be performed by correlation coefficient-based method first, and then through some stochastic methods, such as genetic algorithms (GA), simulated annealing (SA), and ensemble methods. However, different stochastic approaches should be applied simultaneously to obtain essential descriptors influencing the metabolic parameters.

Based on different and appropriate descriptors, QSAR models exploiting from simple multiple regression analysis (MLR) to most modern and complex multivariate analysis or machine-learning methods [44]. The most modern and multivariate approaches are: (i) linear methods such as Partial Least Square (PLS) [45], Linear Discriminant Analysis (LDA) [46, 47], and non-linear statistical methods such as Artificial Neural Networks (ANN) [48], Genetic Algorithm (GA) [49], Support Vector Machines (SVM) [50], Inductive Logic Programming (ILP) [51, 52], k- Nearest Neighbor (kNN) Method [53, 54], Bayesian Modeling [55], Self-Organizing Map (SOM) [50], Multivariate adaptive regression splines (MARS) [56] and graph machines [50]. However, due to the complexity of drug metabolism, different non-linear methods should be chosen concurrently as a modeling approach.

Recent Advances in QSAR Models for CYP450-Mediated Drug Metabolism

QSAR has a long history in the drug discovery field, and reached a tremendous impact in the optimization of promising leads that act on specific targets. Regarding metabolism studies, QSAR models can be used for predicting *in vitro* metabolic stability, for CYP450 inhibition identification and for CYP450 isoform specificity [12, 36, 57-61]. QSAR models for early identification of the predominant CYP450 isoforms responsible for drug metabolism and the specific sites of certain metabolic reactions, not only contribute to the elucidation of drug–drug interactions, but also help to make drug design more predictable and rational in the early stages of drug discovery process [62].

Although CYP450 enzymes represent the main gateway for xenobiotics into human Phase I metabolism, there are many other enzymes that catalyze Phase II metabolism reactions. The enzymes involved at this stage are predominantly transferases with a broad spectrum of substrate specificity, like UGT. Importantly, UGT enzymes also participate in metabolic pathways of many endogenous compounds including steroid hormones, retinoids, and bile acids [63-65].

The successful implementation of *in silico* approaches for modeling CYP450-mediated metabolism has progressed in parallel with the increasing availability of CYP isoform substrate and inhibitor selectivities along with the expanding X-ray crystal structures of human CYP's and homology models. However, the development of models for predicting metabolism and for characterizing structural features of substrates for UGT isoforms is less advanced relative to CYP [66].

Terfloth *et al.* investigated the issue of predicting the isoform specificity for cytochrome P450 3A4, 2D6, and 2C9 substrates [67]. They used 146 compounds for training their models, which were developed using multinomial logistic regression, decision tree, or support vector machine (SVM). From the 146 molecules studied, 126 were reported in DrugBank 2.5 [68] and 63 molecules were metabolized by more than one isoform. More recently, Mishra and co-workers [69] developed SVM based QSAR models to predict substrate specificity of five major isoforms CYP 3A4, 2D6, 1A2, 2C9 and 2C19 of a larger data set of 216 drug molecules created from DrugBank 2.5 [68]. Moreover, the authors have developed a web server *MetaPred* for predicting metabolizing isoforms for a drug molecule [69].

Recently, Freitas *et al.* reported a new method for the identification and separation among substrates and nonsubstrates for a particular CYP isoform [70]. They used a database of 596 substrates of the three most important CYP enzymes (CYP2C9, CYP2D6, and CYP3A4), and 2D and 3D-similarity searches in the determination of the CYP enzyme predominantly responsible for the metabolism of a compound.

COMBINED QUANTUM MECHANICS/MOLECULAR MECHANICS (QM/MM) METHODS

In recent years, driven by the development of new software and advances in hardware technology, it has become evident that the incorporation of quantum mechanical (QM) methods in combination with standard classical approaches, in certain stages of *in silico* drug metabolism studies [7, 16], leads to many improvements. Computational modeling is an essential field for understanding the biological catalysts, reaction intermediates and unstable transition states are crucial to questions of reactivity. Enzymes are large molecules, which means that modeling the reactions that they catalyze is complex and challenging.

Standard molecular mechanics (MM) force fields concerns the treatment of electrostatic effects, which are simply defined by Coulombic interactions between static charges, transferable from system to system. These force fields have been provided good description of protein structure and dynamics, but they cannot be used to model chemical reactions. Molecular dynamics simulations are very important in simulations of protein folding and unfolding [71], in drug design applications [72], and particularly in studies of protein conformational changes [73, 74], simulations of the structure and function of other membrane proteins [75, 76].

Quantum mechanical (QM) methods aim to treat the fundamental quantum mechanics of electronic structure, and so can be used to model chemical reactions. Such quantum chemical methods are more flexible and more generally applicable than molecular mechanics methods. However, QM methods are restricted to systems of up to a few hundred atoms. Thus, the major problem with electronic structure calculations on enzymes is presented by the very large computational resources, which significantly limits the size of the system that can be treated. To overcome this problem,

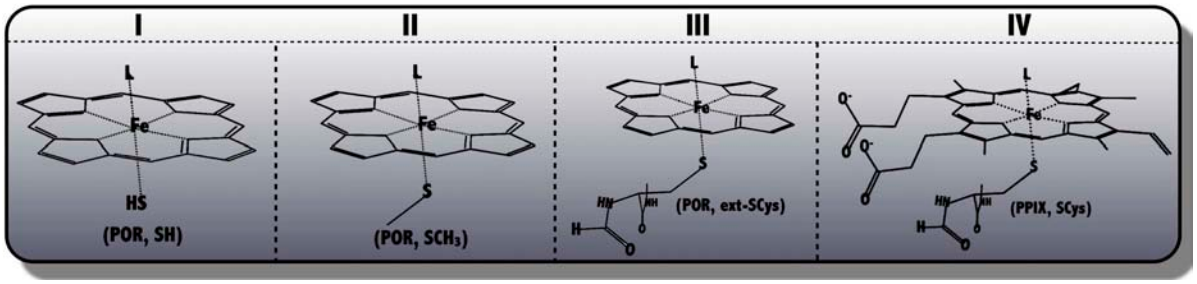

Fig. (3). Different models systems and their notations employed for modeling of CYP450.

small models of enzyme active sites can be studied in isolation. Alternatively, a QM treatment of the chemically active region (e.g. enzyme active site, substrates and co-factors) can be combined with a MM description of the surroundings (e.g. protein and solvent environment): the combined or hybrid QM/MM approach. This methodology will be described below.

QM/MM Approaches for Modeling Cytochrome P450 Reactions

The active species in all CYP-mediated reactions is generally assumed to be a high valent iron(IV)-oxo heme(+•) derivative of the active site heme group known as Compound I (Cpd I) [77]. In Cpd I, the iron is present in an oxidized, oxyferryl (Fe(IV)) form with a triplet spin state, and the porphyrin ring is oxidized to a π-cation radical. In its electronic ground state, Cpd I has two unpaired electrons located in π* orbitals on the Fe-O moiety coupling to one unpaired electron in a π-orbital of the porphyrin ligand with approximate a_{2u} symmetry [16, 78].

One approach to study CYP450 reactions is to analysis just the relatively small active site region. These reduced CYP450 models are shown at Fig. (**3**), allowing the use of powerful QM methods for representing the active site. These range from the simplest models, comprising the heme group without substituents on the porphyrin ring, the SH or SCH_2 group of the cysteinyl ligand, (models I and II), to more realistic ones that involve the axial ligand (cysteinato) (model III) and also vinyl substituents at the β -pyrrole positions (as in the native protoporphyrin IX) (model IV).

The properly *ab initio* methods allow calculations of rate constants for reactions involving very few atoms with results comparable to experimental ones [79]. Semi-empirical molecular orbital techniques such as AM1 [80] and PM3 [81] can model larger systems using linear-scaling methods, allowing performing these calculations on whole proteins. However, semi-empirical methods are well known to frequently give errors of 10 $kcal.mol^{-1}$ or more for calculated barriers and reaction energies [14]. Density functional theory (DFT) methods, especially with the B3LYP hybrid functional, are generally considerably more accurate than semi-empirical methods, and permit calculations on relatively large systems, particularly metalloenzymes such as CYP450 enzymes [14, 78, 82-84].

One notable method is the empirical valence bond (EVB) model, which can be considered as a mixture of force fields of reactant and (intermediates) products in a way that the charge distribution retains the correct variation of the structure along the reaction coordinate. The prominent reliability of the EVB is that the Hamiltonian is calibrated on the reference solution reaction to reproduce experimental (or *ab initio* quantum chemical) in the enzyme active site [85]. These studies clarified the relationship between reactions in solution and enzymes, establishing the catalytic role of preorganized active sites. The EVB method could have a useful application in the late stages of Computer-Aided Enzyme Design (CAED) [86] and in certain stages of *in silico* drug metabolism prediction [7, 16]. Recently, Warshel *et al*. demonstrated the EVB method to be an accurate and reasonably fast method for calculating transition state free energies in different chorismate mutase enzymes and their mutants [87].

The valence bond (VB) diagram model, originally developed for organic reactions, is a theoretical framework that can be used to guide us in the field of bioinorganic chemistry reactivity. The VB diagram model leads to understanding of complex bioinorganic transformations, creates order in the facts, provides an important scaffold for making useful predictions of cytochrome P450 reactivity, and we can rationalize the mechanism during the reactions with the active species of P450, the Cpd I, showed in Fig. (**4**) [4, 78, 83, 88-91].

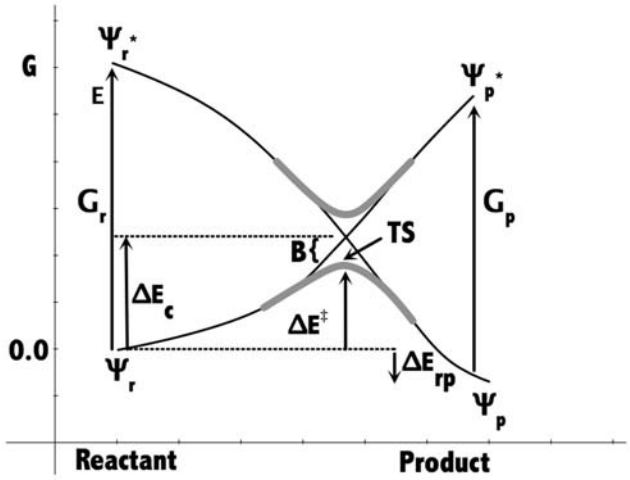

Fig. (4). VB diagram describing the barrier (Δ*E*) formation in an elementary step. G's are promotion energies, B is the resonance energy of the TS, ΔE_{rp} is the reaction energy and the curve describes the result of VB mixing and avoided crossing.

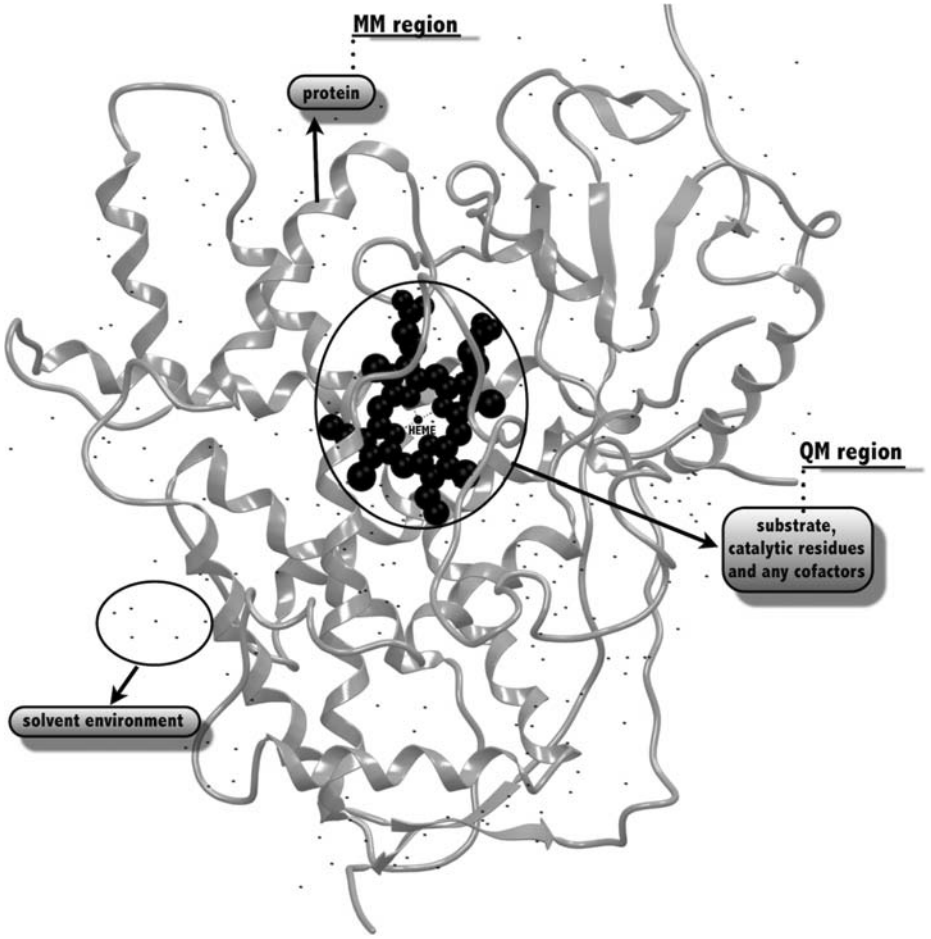

Fig. (5). QM/MM methods are a powerful approach to investigate enzyme mechanisms, specificity and catalysis. The essence of the QM/MM technique is simple: a small region at the active site containing the substrate, catalytic residues and any cofactors (colored in black) is treated by a quantum-chemical method capable of modeling the making and breaking of bonds. This small region interacts with the protein (in gray) and solvent environment (in small black dots), which are treated by a standard empirical molecular-mechanics force field.

Combined quantum-mechanics/molecular-mechanics (QM/MM) approaches have become the method of choice for modeling reactions in biomolecular systems. In simple terms, QM/MM consists of partitioning the system into two domains. A small portion of the macromolecular system (e.g. ligand, or ligand plus its interface with the protein) is treated quantum mechanically using density-based or wavefunction-based methods. However, the size and conformational complexity of the proteins calls for methods capable of treating up to several thousands atoms. This is achieved by using an empirical molecular mechanics force field [92-97]. The resulting schemes are commonly referred to as combined or hybrid QM/MM methods (Fig. **5**). They enable the modeling of reactive biomolecular systems at a reasonable computational effort while providing the necessary accuracy.

Different types of coupling between the QM and MM regions are possible. For applications to CYP450, which is polar, it is important to include the interactions between the QM and MM regions. Modern molecular mechanics methods give a good description of protein structure and interactions ensuring that these are treated accurately. QM/MM calculations can be carried out at semiempirical molecular orbital [98], *ab initio* [99], density-functional [100] or approximate density functional levels [101] of QM electronic structure calculation.

The simplest linking of QM and MM methods involves a straightforward mechanical embedding of the QM region in the MM environment, where the interactions between the QM and MM regions are treated purely classically by MM. In calculations of this type, the QM/MM energy of the whole system, $E_{\mathrm{TOTAL}}^{\mathrm{QM/MM}}$, is calculated in a simple subtractive equation (Eq. 1):

$$E_{\mathrm{TOTAL}}^{\mathrm{QM/MM}} = E_{\mathrm{TOTAL}}^{\mathrm{MM}} + E_{\mathrm{QM\ region}}^{\mathrm{QM}} - E_{\mathrm{QM\ region}}^{\mathrm{MM}} \qquad \text{Eq. 1}$$

Where $E_{\mathrm{TOTAL}}^{\mathrm{MM}}$ is the MM energy of the whole system, $E_{\mathrm{QM\ region}}^{\mathrm{QM}}$, is the QM energy of the QM region and $E_{\mathrm{QM\ region}}^{\mathrm{MM}}$ is the MM energy of the isolated QM region. This subtractive approach can be applied to all combinations of theory levels.

The rapid development of QM/MM in the past few years is making possible to perform calculations to predict activation barriers for CYP450 reactions with accuracy near to 1 kcal/mol [102-104]. At this level, quantitative and reliable predictions can be made about the mechanisms of enzyme-catalysed reactions. This development signals a new era of computational drug metabolism prediction.

Examples of Recent QM/MM Studies of Cytochrome P450 Metabolism

Different CYPP450 isoenzymes show very different substrate specificity and oxidation patterns. These could be the result of orientation or binding effects [105], or the intrinsic chemical reactivity of different positions in the substrates [104]. Genetic polymorphisms can also have significant effects, e.g. in determining drug metabolism [106]. It is possible also that the electronic properties of Compound I could be modulated by the protein environment [90], and that this could be a key factor in determining the reactivity of cytochrome P450s. To investigate these questions, QM/MM calculations that include the protein explicitly are needed. QM/MM modeling of human CYP450 enzymes has been demonstrating the potential of QM/MM methods to contribute directly to practical questions of drug metabolism [82, 102].

The first QM/MM study of human cytochrome P450 enzymes in complex with the drugs diclofenac and ibuprofen has been published in 2005 [16]. The electronic and geometric structure of Compound I was studied with QM/MM calculations. Three human CYPs that are important in drug metabolism (CYP450 2C9, 2B4, and 3B4) were studied. The results showed that Compound I is remarkably similar in all the different P450 enzymes. Substrate complexes were also studied, and it was found that the presence of drug molecules also has essentially no effect on this result. These results indicate that the electronic properties of Compound I in the different human CYP450s are not distinguishable, which implies that observed differences in substrate selectivity are not caused by differences in their electronic properties [16].

Bathelt *et al.* [102] have modeled the hydroxylation of benzene in the enzyme environment of CYP2C9, using QM/MM methods. In contrast to the gas-phase model calculations, the side-on and face-on pathways were found to have similar barriers, indicating that they might compete in reaction. The calculated QM/MM barriers were found to be consistent with the experimental rate constant for benzene hydroxylation in CYP2E1. The rearrangement pathways from the initial s-complex were modeled, to form the epoxide, ketone and *N*-protonated porphyrin species. Epoxide and ketone products were formed with ease in the face on pathway, whereas the epoxide product was favored in the side on pathway. The results conclude that several pathways are energetically possible during P450 mediated aromatic hydroxylation.

INTEGRATING DIFFERENT LEVELS OF THEORY TO STUDY CYTOCHROME P450 METABOLISM

The process of the metabolic reaction of xenobiotic consists of a series of processes, including substrate binding to the enzyme, catalytic reaction of a substrate by the enzyme, and release of a metabolite from the enzyme. At first, the substrate must bind in close proximity between the metabolic reaction atom within the substrate and the catalytic site of the CYP enzyme (*i.e.*, heme oxygen). Force field based docking techniques and molecular dynamics (MD) simulations can mimic this complex formation process and the dynamic motion of the substrate-enzyme complex [107], and therefore, calculate the energies of binding and orientation. Substrate orientation within the active site of CYP450s is a crucial factor for CYP-mediated metabolism. Therefore, docking studies can be particularly useful for gaining selectivity and steric information about potential compounds, which can be used to predict their sites of metabolism and possible toxic metabolites.

MD can simulate the flexibility of CYP active site residues in a time scale, generally in the order of nanoseconds. Depending on the size of the binding pocket, multiple binding poses are possible to give rise to different metabolic products. This might help explain why regioselectivity or stereoselectivity often occurs for CYP450 substrates. With MD simulations combined, docking methods can also be enhanced [11]. QM/MM hybrid method is ideal to determine the orientation and the oxidation energy of a substrate in a CYP450-mediated catalysis. From the calculated energy barrier value, we can tell the absolute or relative oxidation potential in xenobiotic bioactivation. Moreover, at this stage is important to identify the active site residues that could potentially position the substrate for metabolism and stabilize transition states (TS). Fig. (**6**) shows a schematic diagram of a proposed method aiming at improving drug metabolism studies using different levels of theory.

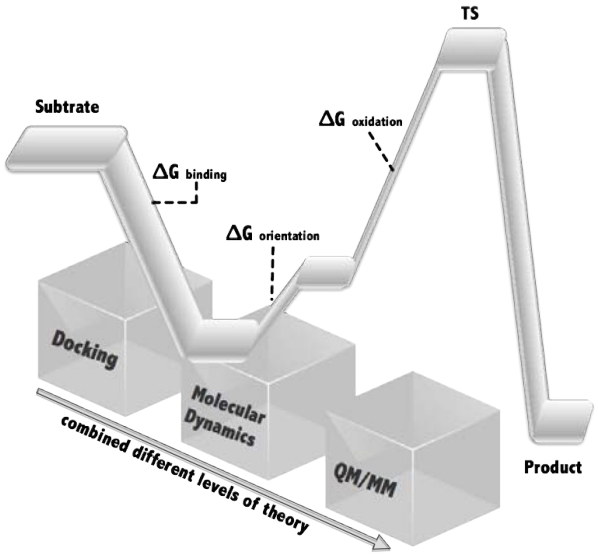

Fig. (6). Proposed method to improve drug metabolism studies using combined different levels of theory, showing the energy changes from substrate binding to product formation in CYP450-catalyzed drug metabolism.

Fig (**6**). also shows the energy changes from substrate binding to product formation in CYP450-catalyzed drug metabolism. Therefore, an integration of computational methods such as QSAR, docking, molecular dynamics and QM/MM calculations can bring us closer to understand drug metabolism and predict drug–drug interactions [11, 107, 108].

CONCLUSIONS

This review has highlighted some aspects that should be emphasized in the QSAR process for drug metabolism, such

as the representation and diversity of the data sets (training and test sets), variable selection and the potential application of novel statistical methods. However, since drug metabolism is an extremely complex pharmacokinetic process, accurate modeling of the drug–metabolic enzyme interactions, including the metabolic degree (hepatic metabolic clearance), type of metabolic enzymes (CYPs or UGTs), type of interactions (substrates, inducers, or inhibitors), site of the interaction (hydrophilic domain or hydrophobic region), and the stereochemical selectivity, is difficult. Also, various approaches should be combined to predict the complex drug metabolism process. For example, QSAR models should be combined with pharmacophore-based approaches or docking methods. As a consequence, the quantitative relationship contained in the QSAR model can be clearly explained and certain mechanism pharmacophore-based or docking methods can also be used to assist in the design of new drugs. Quantum mechanical calculation is a major tool for predicting CYP450 catalysis. From the calculated energy barrier value, we can tell the absolute or relative oxidation potential in xenobiotic metabolism. The identification of the active oxidant in the reaction process is fundamental to understand the formation of products catalyzed by cytochromes P450. Moreover, the integration of *in silico* methods based on a combination of QSAR, docking, molecular dynamics and QM/MM calculations can bring us closer to understand drug metabolism and predict drug–drug interactions.

Although there are many fundamental aspects to be further explored with QSAR and QM/MM techniques, what is clear is that these and other advances will continue to enable and expand the application of these approaches in the metabolism studies for the development of new drugs candidates as an essential ingredient of drug design, and it is likely to remain as such for the foreseeable future.

ACKNOWLEDGEMENT

The authors thank Brazilian Funding Agencies CAPES, CNPq and FAPEG for financial support and fellowships.

REFERENCES

[1] van de Waterbeemd, H.; Gifford, E., ADMET in silico modelling: towards prediction paradise? *Nat. Rev. Drug Discovery*, **2003**, *2*, (3), 192-204.

[2] Li, A.P., Screening for human ADME/Tox drug properties in drug discovery. *Drug Discov. Today*, **2001**, *6*, (7), 357-366.

[3] Kumar, G.N.; Surapaneni, S., Role of drug metabolism in drug discovery and development. *Med. Res. Rev.*, **2001**, *21*, (5), 397-411.

[4] de Montellano, P.R.O., Cytochrome P450: Structure, Mechanism, and Biochemistry. **2010**, 664.

[5] Miners, J.O.; Smith, P.A.; Sorich, M.J.; McKinnon, R.A.; Mackenzie, P.I., Predicting human drug glucuronidation parameters: application of *in vitro* and in silico modeling approaches. *Annu. Rev. Pharmacol. Toxicol.*, **2004**, *44*, 1-25.

[6] Wang, J.F.; Chou, K.C., Molecular modeling of cytochrome P450 and drug metabolism. *Curr. Drug Metab.*, **2010**, *11*, (4), 342-346.

[7] Czodrowski, P.; Kriegl, J.M.; Scheuerer, S.; Fox, T., Computational approaches to predict drug metabolism. *Expert Opin. Drug Metab. Toxicol.*, **2009**, *5*, 15-27.

[8] de Groot, M.J., Designing better drugs: predicting cytochrome P450 metabolism. *Drug Discov. Today*, **2006**, *11*, 601-606.

[9] Khan, M.T.H., Predictions of the ADMET properties of candidate drug molecules utilizing different QSAR/QSPR modelling approaches. *Curr. Drug Metab.*, **2010**, *11*, 285-295.

[10] Kirton, S.B.; Murray, C.W.; Verdonk, M.L.; Taylor, R.D., Prediction of binding modes for ligands in the cytochromes p450 and other heme-containing proteins. *Proteins*, **2005**, *58*, (4), 836-844.

[11] Sun, H.; Scott, D.O., Structure-based drug metabolism predictions for drug design. *Chem. Biol. Drug Des.*, **2010**, *75*, 3-17.

[12] Roy, P.P.; Roy, K., Pharmacophore mapping, molecular docking and QSAR studies of structurally diverse compounds as CYP2B6 inhibitors. *Mol. Simul.*, **2010**, *36*, (11), 887-905.

[13] Senn, H.M.; Thiel, W., QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.*, **2009**, *48*, (7), 1198-1229.

[14] Ridder, L.; Mulholland, A.J.; Rietjens, I.M.C.M.; Vervoort, J., A quantum mechanical/molecular mechanical study of the hydroxylation of phenol and halogenated derivatives by phenol hydroxylase. *J. Am. Chem. Soc.*, **2000**, *122*, (36), 8728-8738.

[15] Ranaghan, K.E.; Mulholland, A.J., Investigations of enzyme-catalysed reactions with combined quantum mechanics/molecular mechanics (QM/MM) methods. *Int. Rev. Phys. Chem.*, **2010**, *29*, (1), 65-133.

[16] Bathelt, C.M.; Zurek, J.; Mulholland, A.J.; Harvey, J.N., Electronic structure of compound I in human isoforms of cytochrome P450 from QM/MM modeling. *J. Am. Chem. Soc.*, **2005**, *127*, (37), 12900-12908.

[17] Smith, P.A.; Sorich, M.J.; McKinnon, R.A.; Miners, J.O., Pharmacophore and quantitative structure-activity relationship modeling: complementary approaches for the rationalization and prediction of UDP-glucuronosyltransferase 1A4 substrate selectivity. *J. Med. Chem.*, **2003**, *46*, (9), 1617-1626.

[18] Madden, J.C.; Cronin, M.T., Structure-based methods for the prediction of drug metabolism. *Expert Opin. Drug Metab. Toxicol.*, **2006**, *2*, (4), 545-557.

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

[20] de Groot, M.J.; Ekins, S., Pharmacophore modeling of cytochromes P450. *Adv. Drug Delivery Rev.*, **2002**, *54*, 367-383.

[21] Hansch, C.; Fujita, T., Rho-Sigma-Pi Analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.*, **1964**, *86*, (8), 1616-1626.

[22] Free, S.M., Jr.; Wilson, J.W., A Mathematical Contribution to Structure-Activity Studies. *J. Med. Chem.*, **1964**, *7*, 395-399.

[23] Andrade, C.H.; Pasqualoto, K.F.; Ferreira, E.I.; Hopfinger, A.J., 4D-QSAR: perspectives in drug design. *Molecules*, **2010**, *15*, (5), 3281-3294.

[24] Cohen, N. *Guidebook on molecular modeling in drug design*; Academic Press: London, **1996**.

[25] Kubinyi, H. In *Methods and Principles in Medicinal Chemistry*. Mannhold, R.; Kroogsgard-Larsen, P.; Timmerman, H., Eds.; Wiley-VCH: Weinheim, **1993**; Vol. *1*, p 240.

[26] Terfloth, L. Gasteiger, J.; Thomas Engel, D., Eds.; Wiley-VCH: Weinheim, **2003**, pp 401-437.

[27] Marshall, G.R.; Cramer, R.D., 3rd, Three-dimensional structure-activity relationships. *Trends Pharmacol. Sci.*, **1988**, *9*, (8), 285-289.

[28] Hopfinger, A.; Wang, S.; Tokarski, J.; Jin, B.; Albuquerque, M.; Madhav, P.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997,** *119*, 10509–10524.

[29] Vedani, A.; Briem, H.; Dobler, M.; Dollinger, H.; McMasters, D.R., Multiple-conformation and protonation-state representation in 4D-QSAR: the neurokinin-1 receptor system. *J. Med. Chem.*, **2000**, *43*, (23), 4416-4427.

[30] Esposito, E.X., Hopfinger, A. J., Madura, J. D. In *Chemoinformatics: concepts, methods, and tools for drug discovery*. Bajorath, J., Ed.; Humana Press: Totowa, **2004**; Vol. *275*, pp 131-213.

[31] Santos-Filho, O.A.; Hopfinger, A.J.; Cherkasov, A.; de Alencastro, R.B., The receptor-dependent QSAR paradigm: an overview of the current state of the art. *Medicinal chemistry (Shāriqah (United Arab Emirates))*, **2009**, *5*, (4), 359-366.

[32] Pan, D.; Tseng, Y.; Hopfinger, A.J., Quantitative structure-based design: formalism and application of receptor-dependent RD-4D-QSAR analysis to a set of glucose analogue inhibitors of glycogen

phosphorylase. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, (5), 1591-1607.

[33] Vedani, A.; Dobler, M., 5D-QSAR: the key for simulating induced fit? *J. Med. Chem.*, **2002**, *45*, (11), 2139-2149.

[34] Vedani, A.; Dobler, M.; Lill, M.A., Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J. Med. Chem.*, **2005**, *48*, (11), 3700-3703.

[35] Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*. Wiley-VCH: Weinheim ; Chichester, **2000**.

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

[37] Bugrim, A.; Nikolskaya, T.; Nikolsky, Y., Early prediction of drug metabolism and toxicity: systems biology approach and modeling. *Drug Discov. Today*, **2004**, *9*, (3), 127-135.

[38] Golbraikh, A.; Tropsha, A., Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J. Comput.-Aided Mol. Des.*, **2002**, *16*, (5-6), 357-369.

[39] [39] Dudek, A.Z.; Arodz, T.; Galvez, J., Computational methods in developing quantitative structure-activity relationships (QSAR): a review. *Comb. Chem. High Throughput Screening*, **2006**, *9*, (3), 213-228.

[40] Cramer III, R.D.; Patterson, D.E.; Bunce, J.D., Comparative Molecular Field Analyses (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.*, **1988**, *110*, 5959-5967.

[41] Kearsley, S.K.; Smith, G.M., An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap. *Tetrahedron Comput. Methodol.*, **1990**, *3*, 615-633

[42] Cruciani, G.; Crivori, P.; Carrupt, P.A.; Testa, B., Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. *J. Mol. Struct. Theochem*, **2000**, *503*, (1-2), 17-30.

[43] Crivori, P.; Zamora, I.; Speed, B.; Orrenius, C.; Poggesi, I., Model based on GRID-derived descriptors for estimating CYP3A4 enzyme stability of potential drug candidates. *J. Comput.-Aided Mol. Des.*, **2004**, *18*, (3), 155-166.

[44] Hou, T.; Wang, J.; Zhang, W.; Wang, W.; Xu, X., Recent advances in computational prediction of drug absorption and permeability in drug discovery. *Curr. Med. Chem.*, **2006**, *13*, (22), 2653-2667.

[45] Dunn Iii, W.J.; Scott, D.R.; Glen, W.G., Principal components analysis and partial least squares regression. *Tetrahedron Comput. Methodol.*, **1989**, *2*, (6), 349-376.

[46] de Julian-Ortiz, J.V.; Garcia-Domenech, R.; Galvez, J.; Pogliani, L., Predictability and prediction of lowest observed adverse effect levels in a structurally heterogeneous set of chemicals. *SAR QSAR Environ. Res.*, **2005**, *16*, (3), 263-272.

[47] Prado-Prado, F.J.; Gonzalez-Diaz, H.; de la Vega, O.M.; Ubeira, F.M.; Chou, K.C., Unified QSAR approach to antimicrobials. Part 3: first multi-tasking QSAR model for input-coded prediction, structural back-projection, and complex networks clustering of antiprotozoal compounds. *Bioorg. Med. Chem.*, **2008**, *16*, (11), 5871-5880.

[48] Hemmateenejad, B., Correlation ranking procedure for factor selection in PC-ANN modeling and application to ADMETox evaluation. *Chemom. Intell. Lab. Syst.*, **2005**, *75*, (2), 231-245.

[49] Rogers, D.; Hopfinger, A.J., Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci.*, **1994**, *34*, (4), 854-866.

[50] Goulon, A.; Picot, T.; Duprat, A.; Dreyfus, G., Predicting activities without computing descriptors: graph machines for QSAR. *SAR QSAR Environ. Res.*, **2007**, *18*, (1), 141-153.

[51] Hirst, J.D.; King, R.D.; Sternberg, M.J., Quantitative structure-activity relationships by neural networks and inductive logic programming. II. The inhibition of dihydrofolate reductase by triazines. *J. Comput.-Aided Mol. Des.*, **1994**, *8*, (4), 421-432.

[52] King, R.D.; Muggleton, S.; Lewis, R.A.; Sternberg, M.J., Drug design by machine learning: the use of inductive logic programming to model the structure-activity relationships of trimethoprim analogues binding to dihydrofolate reductase. *Proc. Natl. Acad. Sci. U. S. A.*, **1992**, *89*, (23), 11322-11326.

[53] Zheng, W.; Tropsha, A., Novel variable selection quantitative structure--property relationship approach based on the k-nearest-neighbor principle. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, (1), 185-194.

[54] Yuan, H.; Wang, Y.; Cheng, Y., Local and global quantitative structure-activity relationship modeling and prediction for the baseline toxicity. *J. Chem. Inf. Model.*, **2007**, *47*, (1), 159-169.

[55] Fernandez, M.; Caballero, J., Bayesian-regularized genetic neural networks applied to the modeling of non-peptide antagonists for the human luteinizing hormone-releasing hormone receptor. *J. Mol. Graphics Modell.*, **2006**, *25*, (4), 410-422.

[56] Ren, S.; Kim, H., Comparative assessment of multiresponse regression methods for predicting the mechanisms of toxic action of phenols. *J. Chem. Inf. Comput. Sci.*, **2003**, *43*, (6), 2106-2110.

[57] Obrezanova, O.; Segall, M.D., Gaussian processes for classification: QSAR modeling of ADMET and target activity. *J. Chem. Inf. Model.*, **2010**, *50*, (6), 1053-1061.

[58] Roy, K.; Roy, P.P., Exploring QSAR and QAAR for inhibitors of cytochrome P450 2A6 and 2A5 enzymes using GFA and G/PLS techniques. *Eur. J. Med. Chem.*, **2009**, *44*, (5), 1941-1951.

[59] Gleeson, M.P.; Davis, A.M.; Chohan, K.K.; Paine, S.W.; Boyer, S.; Gavaghan, C.L.; Arnby, C.H.; Kankkonen, C.; Albertson, N., Generation of in-silico cytochrome P450 1A2, 2C9, 2C19, 2D6, and 3A4 inhibition QSAR models. *J. Comput.-Aided Mol. Des.*, **2007**, *21*, (10-11), 559-573.

[60] Hansch, C.; Mekapati, S.B.; Kurup, A.; Verma, R.P., QSAR of cytochrome P450. *Drug Metab. Rev.*, **2004**, *36*, (1), 105-156.

[61] Hansch, C.; Leo, A.; Mekapati, S.B.; Kurup, A., QSAR and ADME. *Bioorg. Med. Chem.*, **2004**, *12*, 3391-3400.

[62] Wienkers, L.C.; Heath, T.G., Predicting *in vivo* drug interactions from *in vitro* drug discovery data. *Nat. Rev. Drug Discov.*, **2005**, *4*, (10), 825-833.

[63] Czernik, P.J.; Little, J.M.; Barone, G.W.; Raufman, J.P.; Radominska-Pandya, A., Glucuronidation of estrogens and retinoic acid and expression of UDP-glucuronosyltransferase 2B7 in human intestinal mucosa. *Drug Metab. Dispos.*, **2000**, *28*, (10), 1210-1216.

[64] Hum, D.W.; Belanger, A.; Levesque, E.; Barbier, O.; Beaulieu, M.; Albert, C.; Vallee, M.; Guillemette, C.; Tchernof, A.; Turgeon, D.; Dubois, S., Characterization of UDP-glucuronosyltransferases active on steroid hormones. *J. Steroid Biochem. Mol. Biol.*, **1999**, *69*, (1-6), 413-423.

[65] Miley, M.J.; Zielinska, A.K.; Keenan, J.E.; Bratton, S.M.; Radominska-Pandya, A.; Redinbo, M.R., Crystal structure of the cofactor-binding domain of the human phase II drug-metabolism enzyme UDP-glucuronosyltransferase 2B7. *J. Mol. Biol.*, **2007**, *369*, (2), 498-511.

[66] Smith, P.A.; Sorich, M.J.; Low, L.S.; McKinnon, R.A.; Miners, J.O., Towards integrated ADME prediction: past, present and future directions for modelling metabolism by UDP-glucuronosyltransferases. *J. Mol. Graph. Model.*, **2004**, *22*, (6), 507-517.

[67] Terfloth, L.; Bienfait, B.; Gasteiger, J., Ligand-Based Models for the Isoform Specificity of Cytochrome P450 3A4, 2D6, and 2C9 Substrates. *J. Chem. Inf. Model.*, **2007**, *47*, (4), 1688-1701.

[68] Wishart, D.S.; Knox, C.; Guo, A.C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M., DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **2008**, *36*, (Database issue), D901-906.

[69] Mishra, N.K.; Agarwal, S.; Raghava, G.P., Prediction of cytochrome P450 isoform responsible for metabolizing a drug molecule. *BMC Pharmacol.*, **2010**, *10*, 8.

[70] Freitas, R.F.; Bauab, R.L.; Montanari, C.A., Novel Application of 2D and 3D-Similarity Searches To Identify Substrates among Cytochrome P450 2C9, 2D6, and 3A4. *J. Chem. Inf. Model.*, **2010**, *50*, (1), 97-109.

[71] Daggett, V.; Fersht, A., The present view of the mechanism of protein folding. *Nat. Rev. Mol. Cell Biol.*, **2003**, *4*, (6), 497-502.

[72] Morra, G.; Meli, M.; Colombo, G., Molecular dynamics simulations of proteins and peptides: From folding to drug design. *Curr. Protein Pept. Sci.*, **2008**, *9*, (2), 181-196.

[73] Yang, A.Y.C.; Mancera, R.L., Molecular dynamics simulations of ligand-induced backbone conformational changes in the binding site of the periplasmic lysine-, arginine-, ornithine-binding protein. *J. Comput.-Aided Mol. Des.*, **2008**, *22*, (11), 799-814.

[74] Woods, C.J.; Ng, M.H.; Johnston, S.; Murdock, S.E.; Wu, B.; Tai, K.; Fangohr, H.; Jeffreys, P.; Cox, S.; Frey, J.G.; Sansom, M.S.P.; Essex, J.W., Grid computing and biomolecular simulation. *Philos T Roy Soc A*, **2005**, *363*, (1833), 2017-2035.

[75] McGeagh, J.D.; Ranaghan, K.E.; Mulholland, A.J., Protein dynamics and enzyme catalysis: Insights from simulations. *Biochim. Biophys. Acta, Proteins Proteomics*, **2011**, *In Press, Corrected Proof*.

[76] Paneth, P.; Dybala-Defratyka, A., Kinetics and Dynamics: From Nano- to Bio-Scale. *Challenges and Advances in Computational Chemistry and Physics*, **2010**, (12), 575.

[77] Shaik, S.; Kumar, D.; de Visser, S.P.; Altun, A.; Thiel, W., Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.*, **2005**, *105*, 2279-2328.

[78] Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W., P450 enzymes: their structure, reactivity, and selectivity-modeled by QM/MM calculations. *Chem. Rev.*, **2010**, *110*, 949-1017.

[79] Siegbahn, P.E.; Borowski, T., Modeling enzymatic reactions involving transition metals. *Acc. Chem. Res.*, **2006**, *39*, (10), 729-738.

[80] Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P., Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, **1985**, *107*, (13), 3902-3909.

[81] Stewart, J.J.P., Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.*, **1989**, *10*, (2), 209-220.

[82] Mulholland, A.J., Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discov. Today*, **2005**, *10*, (20), 1393-1402.

[83] Lonsdale, R.; Ranaghan, K.E.; Mulholland, A.J., Computational enzymology. *Chemical Communication*, **2010**, *46*, (14), 2354-2372.

[84] Menikarachchi, L.C.; Gascon, J.A., QM/MM Approaches in Medicinal Chemistry Research. *Curr. Top. Med. Chem.*, **2010**, *10*, (1), 46-54.

[85] Field, M.J., Simulating enzyme reactions: Challenges and perspectives. *J. Comput. Chem.*, **2002**, *23*, (1), 48-58.

[86] Toscano, M.D.; Woycechowsky, K.J.; Hilvert, D., Minimalist active-site redesign: Teaching old enzymes new tricks. *Angew. Chem., Int. Ed.*, **2007**, *46*, (18), 3212-3236.

[87] Roca, M.; Vardi-Kilshtain, A.; Warshel, A., Toward Accurate Screening in Computer-Aided Enzyme Design. *Biochemistry*, **2009**, *48*, (14), 3046-3056.

[88] Sono, M.; Roach, M.P.; Coulter, E.D.; Dawson, J.H., Heme-containing oxygenases. *Chem. Rev.*, **1996**, *96*, (7), 2841-2887.

[89] de Montellano, P.R.O.; De Voss, J.J., Oxidizing species in the mechanism of cytochrome P450. *Nat. Prod. Rep.*, **2002**, *19*, (4), 477-493.

[90] Harvey, J.N.; Bathelt, C.M.; Mulholland, A.J., QM/MM modeling of compound I active species in cytochrome P450, Cytochrome c Peroxidase, and Ascorbate Peroxidase. *J. Comput. Chem.*, **2006**, *27*, (12), 1352-1362.

[91] Yoshizawa, K., Theoretical study on kinetic isotope effects in the C-H bond activation of alkanes by iron-oxo complexes. *Coord. Chem. Rev.*, **2002**, *226*, (1-2), 251-259.

[92] Eleftheriou, M.; Germain, R.; Royyuru, A.; Zhou, R.H., Thermal denaturing of mutant lysozyme with both OPLSAA and CHARMM force fields. *Biophys. J.*, **2007**, 193A-193A.

[93] Lii, J.H.; Allinger, N.L., Molecular Mechanics - the Mm3 Force-Field for Hydrocarbons .3. The Vanderwaals Potentials and Crystal Data for Aliphatic and Aromatic-Hydrocarbons. *J. Am. Chem. Soc.*, **1989**, *111*, (23), 8576-8582.

[94] Lii, J.H.; Allinger, N.L., Molecular Mechanics - the Mm3 Force-Field for Hydrocarbons .2. Vibrational Frequencies and Thermodynamics. *J. Am. Chem. Soc.*, **1989**, *111*, (23), 8566-8575.

[95] Allinger, N.L.; Yuh, Y.H.; Lii, J.H., Molecular Mechanics - the Mm3 Force-Field for Hydrocarbons .1. *J. Am. Chem. Soc.*, **1989**, *111*, (23), 8551-8566.

[96] Brooks, B.R.; Brooks, C.L.; Mackerell, A.D.; Nilsson, L.; Petrella, R.J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A.R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R.W.; Post, C.B.; Pu, J.Z.; Schaefer, M.; Tidor, B.; Venable, R.M.; Woodcock, H.L.; Wu, X.; Yang, W.; York, D.M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.*, **2009**, *30*, (10), 1545-1614.

[97] Allinger, N.L.; Kok, R.A.; Imam, M.R., Hydrogen-Bonding in Mm2. *J. Comput. Chem.*, **1988**, *9*, (6), 591-595.

[98] Field, M.J.; Bash, P.A.; Karplus, M., A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J. Comput. Chem.*, **1990**, *11*, (6), 700-733.

[99] Woodcock, H.L.; Hodoscek, M.; Sherwood, P.; Lee, Y.S.; Schaefer, H.F.; Brooks, B.R., Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theor. Chem. Acc.*, **2003**, *109*, (3), 140-148.

[100] Lyne, P.D.; Hodoscek, M.; Karplus, M., A hybrid QM-MM potential employing Hartree-Fock or density functional methods in the quantum region. *J. Phys. Chem. A*, **1999**, *103*, (18), 3462-3471.

[101] Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M., A QM/MM implementation of the self-consistent charge density functional tight binding (SCC-DFTB) method. *J. Phys. Chem. B*, **2001**, *105*, (2), 569-585.

[102] Bathelt, C.M.; Mulholland, A.J.; Harvey, J.N., QM/MM modeling of benzene hydroxylation in human cytochrome P450 2C9. *J. Phys. Chem. A*, **2008**, *112*, (50), 13149-13156.

[103] Lonsdale, R.; Harvey, J.N.; Mulholland, A.J., Inclusion of Dispersion Effects Significantly Improves Accuracy of Calculated Reaction Barriers for Cytochrome P450 Catalyzed Reactions. *J. Phys. Chem. Lett.*, **2010**, *1*, (21), 3232-3237.

[104] Harvey, J.N.; Aggarwal, V.K.; Bathelt, C.M.; Carreón-Macedo, J.-L.; Gallagher, T.; Holzmann, N.; Mulholland, A.J.; Robiette, R., QM and QM/MM studies of selectivity in organic and bioorganic chemistry. *J. Phys. Org. Chem.*, **2006**, *19*, (8-9), 608-615.

[105] de Groot, M.J.; Kirton, S.B.; Sutcliffe, M.J., In silico methods for predicting ligand binding determinants of cytochromes P450. *Curr. Top. Med. Chem.*, **2004**, *4*, (16), 1803-1824.

[106] Pirmohamed, M.; Park, B.K., Cytochrome P450 enzyme polymorphisms and adverse drug reactions. *Toxicology*, **2003**, *192*, (1), 23-32.

[107] Park, J.-Y.; Harris, D., Construction and assessment of models of CYP2E1: predictions of metabolism from docking, molecular dynamics, and density functional theoretical calculations. *J. Med. Chem.*, **2003**, *46*, (9), 1645-1660.

[108] de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N.P.E., Binding mode prediction of cytochrome p450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem.*, **2005**, *48*, (7), 2308-2318.

Received: March 16, 2011 Revised: September 30, 2011 Accepted: October 3, 2011