ENVIRONMENTAL TOXICOLOGY AND PHARMACOLOGY 39 (2015) 247–258

Available online at www.sciencedirect.com

ScienceDirect

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

# The utilisation of structural descriptors to predict metabolic constants of xenobiotics in mammals

*Alessandra Pirovano*[a,*], *Stefan Brandmaier*[b,c], *Mark A.J. Huijbregts*[a], *Ad M.J. Ragas*[a,d], *Karin Veltman*[a], *A. Jan Hendriks*[a]

[a] *Radboud University Nijmegen, Institute for Wetland and Water Research, Department of Environmental Science, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands*
[b] *Helmholtz Zentrum München—German Research Centre for Environmental Health (GmbH), Institute of Structural Biology, Ingolstaedter Landstrasse 1, Neuherberg, D-85764 Munich, Germany*
[c] *Helmholtz-Zentrum München—German Research Centre for Environmental Health (GmbH), Research Unit of Molecular Epidemiology, Institute of Epidemiology II, Munich, Germany*
[d] *Faculty of Management, Science and Technology, Open University, Heerlen, The Netherlands*

ARTICLE INFO

*Article history:*
Received 8 July 2014
Received in revised form
28 November 2014
Accepted 29 November 2014
Available online 5 December 2014

*Keywords:*
Quantitative structure activity relationship
Biotransformation
Michaelis–Menten constant ($K_m$)
Maximum reaction rate ($V_{max}$)

ABSTRACT

Quantitative structure–activity relationships (QSARs) were developed to predict the Michaelis–Menten constant ($K_m$) and the maximum reaction rate ($V_{max}$) of xenobiotics metabolised by four enzyme classes in mammalian livers: alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), flavin-containing monooxygenase (FMO), and cytochrome P450 (CYP). Metabolic constants were gathered from the literature and a genetic algorithm was employed to select at most six predictors from a pool of over 2000 potential molecular descriptors using two-thirds of the xenobiotics in each enzyme class. The resulting multiple linear models were cross-validated using the remaining one-third of the compounds. The explained variances ($R^2_{adj}$) of the QSARs were between 50% and 80% and the predictive abilities ($R^2_{ext}$) between 50% and 60%, except for the $V_{max}$ QSAR of FMO with both $R^2_{adj}$ and $R^2_{ext}$ less than 30%. The $V_{max}$ values of FMO were independent of substrate chemical structure because the rate-limiting step of its catalytic cycle occurs before compound oxidation. For the other enzymes, $V_{max}$ was predominantly determined by functional groups or fragments and electronic properties because of the strong and chemical-specific interactions involved in the metabolic reactions. The most relevant predictors for $K_m$ were functional groups or fragments for the enzymes metabolising specific compounds (ADH, ALDH and FMO) and size and shape properties for CYP, likely because of the broad substrate specificity of CYP enzymes. The present study can be helpful to predict the $K_m$ and $V_{max}$ of four important oxidising enzymes in mammals and better understand the underlying principles of chemical transformation by liver enzymes.



* Corresponding author. Tel.: +31 243652725.
E-mail address: a.pirovano@science.ru.nl (A. Pirovano).

http://dx.doi.org/10.1016/j.etap.2014.11.025
1382-6689/

## 1. Introduction

Information regarding the biotransformation of xenobiotics is essential for environmental toxicology, risk assessment and drug development because metabolism can largely influence the residence time and bioaccumulation of chemicals in organisms (Sijm et al., 2007; Wang et al., 2014). Through biotransformation, the parent compound (substrate) is converted by enzymes into another chemical (metabolite), which is usually more soluble and thus can be excreted more easily. Metabolism occurs via enzymatic reactions involving two processes. First, the chemical needs to reach the enzyme and bind with it; second, a catalytic reaction must occur. The latter process is described by the maximum rate of reaction ($V_{max}$) at saturating substrate concentration (Testa et al., 2000). The other parameter used to characterise an enzymatic reaction is the Michaelis–Menten constant ($K_m$), which is the substrate concentration at half the maximum rate, i.e., at $V_{max}/2$. If the catalytic step is slow compared with the dissociation of the substrate from the enzyme, $K_m$ is assumed to be equal to the dissociation constant $K_d$ for the enzyme-substrate complex. In this case, the inverse of the Michaelis–Menten constant ($1/K_m$) reflects the affinity of the enzyme for its substrate: a high $1/K_m$ corresponds to high binding affinity. For reactions that exhibit Michaelis–Menten kinetics and at non-saturating substrate concentrations, the ratio between $V_{max}$ and $K_m$ estimates intrinsic clearance ($CL_{int}$). Intrinsic clearance, which is a measure of enzyme activity towards a compound, can be extrapolated to an equivalent whole-body metabolic rate required for risk assessment (Blaauboer, 2002; Lipscomb and Poet, 2008).

Measured $K_m$ and $V_{max}$ values are lacking for many chemicals and species. In silico methods, such as quantitative structure activity relationships (QSARs), can be useful tools for predicting biological transformation rates on the basis of chemical descriptors (Cherkasov et al., 2013). In previous studies, metabolic constants were frequently found to correlate with easily interpretable physico-chemical properties of substrates, such as hydrophobicity or hydrogen bonding (Balaz, 2009). However, the reported QSARs had generally low explained variances (Pirovano et al., 2014) or considered only a limited series of substrates (Hansch et al., 2004). Weak correlations indicated that the metabolic processes could only partly be explained by the physico-chemical descriptors chosen, possibly because of the complexity of the underlying metabolic reactions (Waller et al., 1996). In the present study, we included a large number of theoretical molecular descriptors (approximately 2000), such as topological indices and functional group counts, which can capture the structural and molecular information of chemicals (Consonni and Todeschini, 2010). The use of theoretical molecular descriptors in QSAR models is helpful to identify the chemical features influencing the biological activities of large sets of diverse chemicals.

The aim of this study was to develop QSARs for the affinity constant ($1/K_m$) and maximum reaction rate of xenobiotics transformed by the alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), flavin-containing monooxygenase (FMO) and cytochrome P450 (CYP) enzymes in mammals. The QSARs were built with multiple linear regressions (MLR) by selecting theoretical descriptors with genetic algorithms. The QSARs were mechanistically interpreted to provide insight into the processes governing biotransformation. External validation was applied to assess the predictive power of the models.

## 2. Materials and methods

### 2.1. Experimental dataset

The enzymatic constants ($K_m$ and $V_{max}$) were collected from the scientific literature for alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), flavin-containing monooxygenase (FMO), and cytochrome P450 (CYP). Liver ADH catalyses the reversible transformation of alcohols to their corresponding aldehydes or ketones. ALDH enzymes oxidise a wide range of aldehydes to their corresponding carboxylic acids (Vasiliou et al., 2000). FMO oxygenates a wide range of xenobiotics that contain a nucleophilic heteroatom (usually sulphur and nitrogen, with the oxidative reaction resulting in the formation of N or S-oxides), such as pesticides and drugs (Krueger and Williams, 2005). P450 enzymes usually catalyse monooxygenase reactions, which involve the insertion of an oxygen atom into a substrate (Lewis, 1999).

Data were taken from the BRENDA enzyme database (Scheer et al., 2011) and several reviews (Hansch et al., 2004; Hansch and Zhang, 1993; Krueger and Williams, 2005; Lewis, 1999). Constants measured for mammals in in vitro assays of purified, non-recombinant, hepatic enzymes were selected. Data were available for different isoenzymes (i.e., any of the several forms of an enzyme, all of which catalyse the same reaction but are characterised by different properties) and for the following species: horse (ADH, ALDH), human (ADH, ALDH), rat (ADH, ALDH, CYP), mouse (FMO), pig (FMO), and rabbit (CYP). All data were checked in the original papers and are reported in the Supporting Information (SI, Table S1).

$K_m$ values were expressed in μM and all rates were expressed as $V_{max}$ with $\mu mol\,min^{-1}\,mg_{PROT}{}^{-1}$ as units. The rates were reported in the papers either as $V_{max}$ or as catalytic constant ($k_{cat}$) values. The latter were transformed into $V_{max}$ using the weight of the enzyme or the content of microsomal protein (for CYP) as conversion factors. We used the values reported in the studies measuring $k_{cat}$, when reported; otherwise, we used the average values obtained from other studies (Table S2 in the SI).

$K_m$ and $V_{max}$ data were combined into four databases, one for each enzyme family, independently of the species and isoenzyme. Each substrate was characterised by a single value of $1/K_m$ or $V_{max}$. If multiple values were available for one substrate, we calculated the geometric mean and standard deviation of the experimental $1/K_m$ or $V_{max}$ values. The compounds collected were represented as SMILES (simplified molecular input line entry system) strings.

### 2.2. Molecular descriptors

Approximately 2000 descriptors were calculated using the Online CHEmical Modeling environment platform (OCHEM) (Sushko et al., 2011). These descriptors included