

## **Quantifying biotransformation of xenobiotics in mammals**

Pirovano A., 2015. Quantifying biotransformation of xenobiotics in mammals.  
PhD thesis, Radboud University Nijmegen, the Netherlands.

©2015 Alessandra Pirovano, all rights reserved.

ISBN: 978-90-6464-939-4

Printed by: GVO Drukkers & Vormgevers, Ede, The Netherlands

Cover picture source: [colourlovers.com](http://colourlovers.com)

# **Quantifying biotransformation of xenobiotics in mammals**

Proefschrift

ter verkrijging van de graad van doctor  
aan de Radboud Universiteit Nijmegen  
op gezag van de rector magnificus,  
volgens besluit van het college van decanen  
in het openbaar te verdedigen op woensdag 9 december 2015  
om 10:00 uur precies

door

**Alessandra Pirovano**

geboren op 16 april 1984  
te Melzo (Milano), Italië

**Promotoren:**

Prof. dr. ir. A.J. Hendriks

Prof. dr. M.A.J. Huijbregts

Prof. dr. A.M.J. Ragas (Open Universiteit)

**Copromotor:**

Dr. K. Veltman (University of Michigan, Verenigde Staten)

**Manuscriptcommissie:**

Prof. dr. A.M. Breure

Prof. dr. F.G.M. Russel

Prof. dr. B.J. Blaauboer (Universiteit Utrecht)

The research described in this thesis was performed within project ECO (Environmental ChemOinformatics). The project was financially supported by the European Union (FP7-PEOPLE-ITN-2008, n. 238701).



# Table of contents

Chapter 1	
General Introduction .....	7
Chapter 2	
A comparison of octanol-water partitioning between organic chemicals and their metabolites in mammals .....	17
Chapter 3	
Compound lipophilicity as a descriptor to predict binding affinity ( $1/K_m$ ) in mammals .....	29
Chapter 4	
Mechanistically-based QSARs to describe metabolic constants in mammals ..	45
Chapter 5	
The utilisation of structural descriptors to predict metabolic constants of xenobiotics in mammals.....	63
Chapter 6	
QSARs for estimating intrinsic hepatic clearance of organic chemicals in humans .....	85
Chapter 7	
Synthesis.....	101
Appendix A .....	115
Appendix B .....	135
Appendix C .....	189
Appendix D .....	201
Appendix E.....	207
Appendix F .....	211
Appendix G .....	225
References.....	231
Summary .....	263
Samenvatting.....	266
About the author.....	269
Acknowledgments.....	270



# Chapter 1

## General Introduction

## 1.1 Background

### 1.1.1 Risk assessment of chemicals

Manufactured chemicals are widely used in our society for multiple purposes, such as medical treatment, crop protection and house cleaning. The manufacturing, transfer and use of these products result in the release of thousands of xenobiotics into the environment. After emission, chemicals are transported and distributed across air, water, soil and sediment, where they may be degraded into transformation products by biotic or abiotic processes [1]. These xenobiotics and their transformation products can be taken up by aquatic and terrestrial organisms via different exposure routes, such as inhalation, absorption and food intake. The concentration of chemicals in organisms can be reduced through elimination processes, such as exhalation and egestion, or via biotransformation reactions mediated by enzymes. Chemicals can eventually accumulate in organisms if their uptake rate is faster than their elimination rate from the organism. It is important to determine to what degree chemicals accumulate, since xenobiotics that enter the body may exert hazardous effects on humans and animals.

Regulators have taken measures to improve the protection of organisms and ecosystems against the risks that can be posed by exposure to chemicals. For example, the European Union (EU) adopted the REACH (Registration, Evaluation, Authorization and restriction of Chemicals) legislation, which entered into force on 1 June 2007 [2]. The REACH regulation makes companies responsible to ensure that the substances they manufacture and market in the EU can be used safely. Registrants must provide data on physicochemical and (eco)toxicological properties of substances, following clearly defined information requirements that are tonnage and risk related [3]. These data have to be used to assess the risks arising from the entire chemical life cycle, as well as to develop and recommend appropriate risk management measures to control these risks. The information gathered and the assessment performed must be submitted to the European Chemicals Agency (ECHA) to be evaluated for the registration [4].

The data needed for the risk assessment of chemicals include toxicological and (eco)toxicological endpoints such as skin irritation, mutagenicity, terrestrial and aquatic toxicity, bioaccumulation, etc. These data are conventionally measured in laboratory experiments. Due to ethical, financial and practical constraints, not all chemicals can be tested on all species [5]. Thus, REACH promotes alternative methods to replace, reduce and refine the use of animals in scientific procedures (3Rs principle), provided that the use of reliable alternative methods is justified with a scientific explanation. Alternative estimation methods include *in vitro* experiments [6] and *in silico* models [7].

### 1.1.2 Quantitative Structure Activity Relationship

Quantitative Structure Activity Relationships (QSARs) represent a widely used *in silico* modelling approach for estimating the biological activity of a substance from features of its chemical structure. The fundamental assumption of QSARs is that the structure of a chemical implicitly determines its physicochemical properties, which, in interaction with a biological system, determine its (eco)toxicological properties [8]. QSAR modelling generally involves three steps: 1) collection of experimental data measuring the property or biological activity of interest (endpoint) for different chemicals; 2) calculation of descriptors that represent properties and/or features of the molecular structure of the chemicals; 3) application of statistical methods that relate descriptors to the endpoint. One of the most common and transparent methods is Multiple Linear Regression (MLR), where the endpoint is expressed as a linear function of a limited number of descriptors [9]. The development of QSARs has two main practical purposes. First, it provides insights into mechanisms of biological processes and allows for the identification of important structural characteristics and/or physicochemical properties influencing the endpoint. Second, it allows for the prediction of the biological activity of untested chemicals from their structures, thus contributing to the 3Rs in the risk assessment of chemicals [10].

The appropriate descriptors to model a defined endpoint can be chosen with two main approaches, depending on the aim of the QSAR. In the “mechanistic” approach, chemical structure is represented only by few molecular descriptors of clear physicochemical interpretation, related to the size, chemical reactivity and partitioning of the substances. For example, the octanol-water partition coefficient ( $K_{ow}$ ) has often been related to many different endpoints, e.g. soil sorption, bioaccumulation and baseline toxicity [11], as it approximates the ability of a chemical to reach the site of action. The  $K_{ow}$  is defined as the ratio of the concentration of a chemical in n-octanol and water at equilibrium and represents the hydrophobicity (or lipophilicity) of a compound. The “mechanistic” descriptors are chosen by the modeller on the basis of *a priori* knowledge of the mechanism of the endpoint [12], with the aim to enhance understanding and provide a more rational basis for risk assessment. Alternatively, in the “statistical” approach, chemical structure is represented by a large number (usually thousands) of theoretical molecular descriptors, such as topological and fragment based indices, which encode multiple aspects of the molecular structure. The “theoretical” descriptors for the QSAR are then selected by different chemometric methods as the best correlated with the endpoint, with the main aim to optimise model performance for prediction [12].

## 1.2 Biotransformation of chemicals

### 1.2.1 The role of biotransformation in bioaccumulation modelling

In bioaccumulation modelling, biotransformation is one of the processes that decrease the concentration of metabolisable compounds in an organism, together with elimination through other physiological processes, such as exhalation and egestion. Through biotransformation, the parent compound is converted via enzymatic reactions into another chemical (metabolite), which is usually more soluble and thus can be excreted more easily. There are two types of biotransformation reactions: Phase 1 (hydrolysis, reduction and oxidation) and Phase 2 (conjugation) reactions [13]. During Phase 1 reactions, the parent compound is transformed by introducing polar functional groups (such as -OH, -COOH or -NH<sub>2</sub>). Phase 2 reactions combine the substrate (a parent compound or more commonly a Phase 1 metabolite) with an endogenous substance (such as glutathione, glucuronide or acetic acid). To be metabolised, the chemical must reach the enzyme and bind to it; then, a catalytic reaction must occur. Therefore, the biotransformation rate ( $k_m$ , d<sup>-1</sup>) is determined both by the internal distribution and the capacity of the enzyme to bind and transform the substrate [1].

Models have been developed to assess the bioaccumulation of chemicals by quantifying the kinetic rate constants of uptake and elimination (mass balance models) [14, 15]. Rates of elimination via exhalation with air, excretion with urine and egestion of non-digested food can be predicted quite accurately from properties of chemical substances and biological species, such as chemical Log K<sub>ow</sub> and organism size [15, 16]. On the contrary, biotransformation rates ( $k_m$ , d<sup>-1</sup>) are difficult to estimate because they apply to a specific combination of a chemical and enzymes and vary among individual organisms and species. In fact, multiple enzyme systems exist and the overall metabolic rate depends on the enzyme composition, i.e. concentration and activity. Because of the lack of information regarding biotransformation capabilities,  $k_m$  values were often not considered in the determination of bioaccumulation of chemicals, leading to overestimation of the bioaccumulation for metabolisable chemicals [16]. Biotransformation was in fact shown to largely influence bioaccumulation of metabolisable chemicals in both mammals and fish [17, 18].

### 1.2.2 Quantification of biotransformation

Limited  $k_m$  data measured for the whole-body *in vivo* are available in the scientific literature, since it is difficult to isolate metabolism from the plethora of other physiological processes [19, 20]. Because of the important contribution of biotransformation to the bioaccumulation of chemicals, many

efforts have recently been made to obtain  $k_m$  values following generally two approaches: 1) from measured total elimination rates using (mechanistic) mass-balance models; 2) by extrapolating *in vitro* measurements of the metabolic constants to their whole-body *in vivo* equivalents, as explained in Chapter 7. In the first approach, the biotransformation rate of organic chemicals can be estimated for various species groups as the difference between measured elimination rate constants and the sum of elimination rate constants predicted assuming no metabolism [20]. For humans, biotransformation rates have recently been estimated from measured total elimination rates with a mass balance model [21] and subsequently used to develop QSARs. In the second approach, the biotransformation potential is commonly assayed via the measurement of intrinsic clearance ( $CL_{INT}$ ,  $\text{mL min}^{-1} \text{kg}_{BW}^{-1}$ ) in *in vitro* systems derived from liver tissue, such as isolated hepatocytes, microsomes, S9 fractions or isolated enzymes. Liver is in fact the principal organ responsible for the metabolism in fish and mammals [1, 22]. The *in vitro*  $CL_{INT}$  is calculated as the ratio between the maximum reaction rate ( $V_{max}$ ) and the Michaelis-Menten constant ( $K_m$ ). The hepatic  $CL_{INT}$  is then incorporated into established physiologically based models for the estimation of  $k_m$  values [19]. A stepwise approach for *in vitro* to *in vivo* extrapolation (ivive) was initially developed for mammals by the pharmaceutical industry to support preclinical screening of drug candidates [23].

### 1.2.3 Kinetics of biotransformation

Understanding enzyme kinetics is important to determine the metabolic rate and to obtain a better mechanistic understanding of biotransformation reactions [1]. Enzymes are proteins and their catalytic function occurs within a pocket named active site. The surface of the enzyme active site is lined with functional groups (amino acid side chains, inorganic metal ions or coenzymes) that bind the substrate and then catalyse its chemical transformation into a product, leaving the enzyme chemically unchanged [24]. When the substrate reaches the enzyme, the functional groups on the active site sequester the chemical from aquatic solution, forming a transient enzyme-substrate complex via weak non-covalent interactions (hydrogen bonds, hydrophobic and ionic interactions). The weak binding interactions between enzyme and substrate contribute to its successive catalysis, as they hold the substrate and bring specific functional groups into the optimal position to react. In the catalytic step, the cleavage and formation of covalent or ionic bonds between the substrate and the catalytic functional groups result in the release of the product and the return of the enzyme to its original state.

The enzymatic reaction can be described as follows:

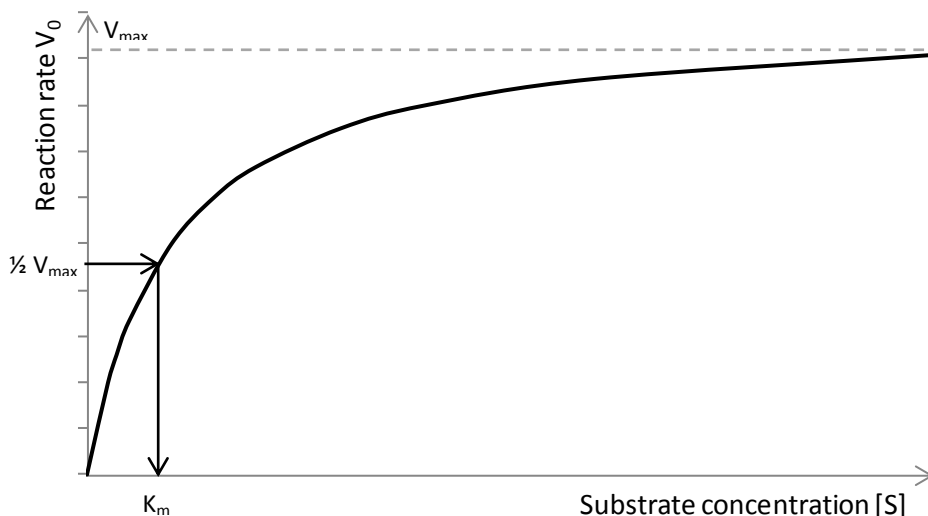


where E, S and P represent the enzyme, substrate and product; ES is the enzyme-substrate complex; and  $k_1$ ,  $k_{-1}$ ,  $k_2$ ,  $k_{-2}$  ( $\text{d}^{-1}$ ) are the rate constants for formation and breakdown of ES. Early in the reaction, product concentration [P] is negligible, thus the reverse reaction  $\text{P} \rightarrow \text{ES}$  described by  $k_{-2}$  is assumed to be negligible. The enzymatic reaction can be rewritten as follows:



where  $k_{cat}$  ( $\text{d}^{-1}$ ) is the rate constant for P formation, which is usually the rate limiting step in the overall enzymatic reaction [24]. The rate constant  $k_{cat}$  is also named turnover number and represents the amount of S converted to P per time unit on a single enzyme molecule. The initial reaction rate ( $V_0$ ,  $\text{mol min}^{-1} \text{mg}_E^{-1}$ ) is defined as the amount of P formed per time unit per amount of enzyme. For many enzymes,  $V_0$  varies with substrate concentration ( $[\text{S}]$ ,  $\text{mol L}^{-1}$ ) following the typical Michaelis-Menten plot shown in Figure 1.1, assuming the total enzyme concentration  $[\text{E}_T]$  to be constant and considerably smaller than  $[\text{S}]$ .

Figure 1.1. Effect of substrate concentration [S] on the initial rate of an enzyme-catalysed reaction ( $V_0$ ).





At lower [S],  $V_0$  increases linearly with substrate concentration. At higher [S],  $V_0$  begins to level off until it approaches a maximum and the reaction is saturated (steady-state). The reaction rate is given by the following equation:

$$V_0 = \frac{V_{\max} \cdot [S]}{K_m + [S]} \quad (\text{Eq. 1.3})$$

Where  $V_{\max}$  ( $\text{mol min}^{-1} \text{mg}_E^{-1}$ ) is the maximum reaction rate and  $K_m$  ( $\text{mol L}^{-1}$ ) is the substrate concentration at half  $V_{\max}$ . The catalytic step of the enzymatic reaction is described by  $V_{\max}$ , which is equal to the product between  $k_{\text{cat}}$  and  $[E_T]$ . The Michaelis-Menten constant  $K_m$  is independent of  $[E_T]$  and typically describes the binding step [25]. If the catalytic step is slow compared with the dissociation of S from E ( $k_{\text{cat}} \ll k_{-1}$ ),  $K_m$  reduces to  $k_{-1}/k_1$ , which is defined as the dissociation constant  $K_d$  of the ES complex. In this case, the inverse of  $K_m$  reflects the affinity of the enzyme for its substrate: a high  $1/K_m$  (or low  $K_m$ ) corresponds to high binding affinity [24].

### 1.3 Problem setting

In environmental modelling, the prediction of the biotransformation rate is a difficult task due to the specific action of metabolism, which depends on the chemical and the enzyme involved and varies among individual organisms and species. Enzymes determine the qualitative and quantitative aspects of biotransformation [1], thus investigations on the mechanisms governing metabolism should start from the enzyme level.

QSARs have been built to estimate the enzymatic constants ( $K_m$  and  $V_{\max}$ ) for drugs oxidised by cytochrome P450 (CYP) in mammals [26, 27]. These constants were correlated with mechanistic descriptors representing easily interpretable physicochemical properties of substrates. The binding affinity, represented by  $1/K_m$ , was mainly correlated with compound hydrophobicity, expressed as  $\text{Log } K_{\text{ow}}$  [25, 28], probably because of desolvation effects. The maximum rate  $V_{\max}$  was mostly influenced by electronic properties, such as frontier orbital energies or hydrogen bonding [29-31]. In fact, catalytic processes are characterised by cleavage and formation of covalent bonds [25]. However, the above-mentioned studies considered only a limited series of P450 substrates, mainly drugs. CYP is the major (and thus the most studied) enzyme group in terms of catalytic versatility and the large number of xenobiotics it detoxifies or activates [13]. Nevertheless, the contribution of other enzymes to the oxidative metabolism of xenobiotics is significant as well [32]. Despite their importance, QSARs for non-CYP enzymes have hardly been developed. In addition, the above mentioned studies only used mechanistic descriptors. Given the complexity of the underlying metabolic reactions,

theoretical molecular descriptors (such as topological indices and functional group counts) might be more appropriate to identify the chemical features influencing metabolism of large sets of diverse chemicals.

In order to quantify biotransformation rates, it is necessary to obtain  $K_m$  and  $V_{max}$  values measured in *in vitro* systems derived from liver tissue (e.g. isolated hepatocytes, microsomes), which have to be extrapolated to their whole-body *in vivo* equivalents. Measurements of  $K_m$  and  $V_{max}$  values from liver tissue are lacking for many chemicals and species. A few models have been built to predict *in vitro* clearance for mammals (measured mainly in microsomes or hepatocytes) using information on the chemical structure [33-36], but these models included only pharmaceuticals. Although data are available, no QSARs have been developed yet to predict *in vitro*  $CL_{INT}$  including environmental pollutants. In addition, the *in vivo* methods developed for mammals were used mainly for drugs, with the aim to accelerate the selection of new candidates in the drug discovery stage based on their predicted clearance. Despite the importance of biotransformation for the risk assessment of environmental pollutants, few attempts have been made to derive  $k_m$  values from *in vivo* methods.

#### 1.4 Aims and outline

The overall aim of this thesis is to develop QSARs for the prediction of biotransformation of xenobiotics in mammals based on their chemical properties.

Compared with previous QSARs for biotransformation that were available only for drugs, the focus of this thesis is on both pharmaceuticals and environmental pollutants metabolised in mammals. In addition, the relationships between metabolic activity and chemical structure were developed using different types of descriptors, first  $K_{ow}$  only, then mechanistic descriptors and finally theoretical descriptors. Moreover, QSARs were developed for systems representing different levels of biological organization (isolated enzymes, hepatocytes and microsomes). In Figure 1.2 a schematic overview is given of the thesis content.

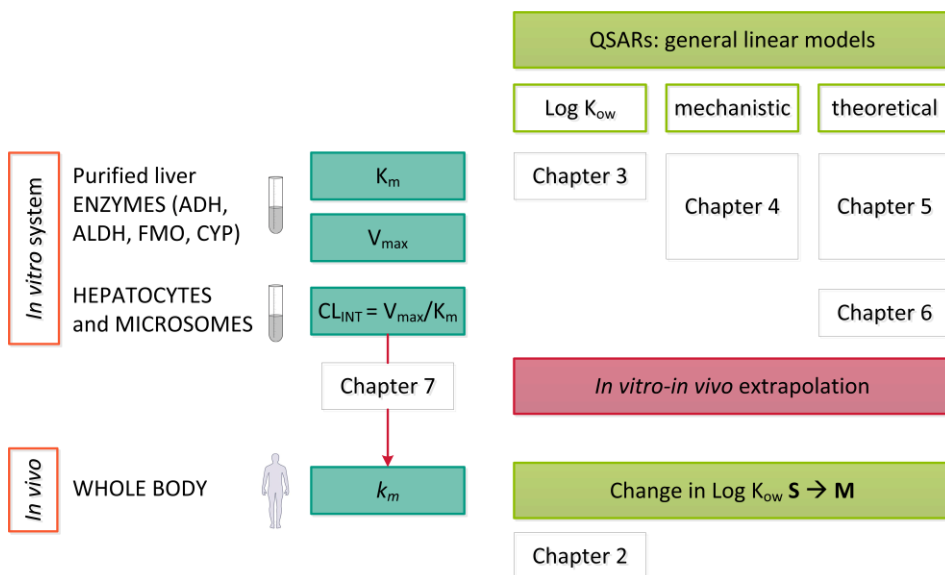
The general mechanisms underlying metabolism were investigated starting from the enzyme level. The focus was on the liver metabolism in mammals mediated by four important oxidising enzymes: ADH, ALDH, FMO and CYP. First, the influence of compound hydrophobicity ( $\log K_{ow}$ ) on metabolism was investigated. In Chapter 2, the change in  $\log K_{ow}$  of the parent compound after it is metabolised was quantified. In Chapter 3, the relationships between  $\log K_{ow}$  and the  $K_m$  values measured in purified enzymes were investigated. Next,

the relationships between the metabolic constants ( $K_m$  and  $V_{max}$  measured in purified enzymes) and chemical properties were analysed. QSARs were developed using mechanistic descriptors known to influence metabolism (Chapter 4), as well as theoretical descriptors (Chapter 5).

Successively,  $K_m$  and  $V_{max}$  values were also collected for whole liver cells and sub-cellular fractions (hepatocytes and microsomes) to build QSARs predicting clearance, i.e.  $V_{max}/K_m$  (Chapter 6). These models were interpreted also in the light of the results found for enzymes.

Finally, in Chapter 7 the advantages and disadvantages of the different types of descriptors and levels of biological organization are discussed. A general scheme was developed to perform *in vitro-in vivo* extrapolations (ivive). This scheme was used to derive  $k_m$  values using clearance collected for human microsomes and hepatocytes. The extrapolated  $k_m$  values were compared to *in vivo* measurements in order to validate the ivive method. Finally, a tentative refinement of the accumulation of the parent compound based on the change in hydrophobicity after metabolism is discussed.

Figure 1.2. Thesis content





## Chapter 2

# **A comparison of octanol-water partitioning between organic chemicals and their metabolites in mammals**

Alessandra Pirovano

Nicolò Borile

A. Jan Hendriks

*Chemosphere* (2012), 88(8), 1036–1041

## 2.1 Introduction

Risk assessment of xenobiotics present in the environment needs comprehensive evaluation of accumulation potential in organisms. Recently developed *in silico* mechanistic models estimate the bioaccumulation factors of chemicals, calculated as the difference between uptake and elimination rates from organisms [15]. In addition to the excretion via urine, egestion via feces and growth dilution, labile compounds can be eliminated by metabolism. Yet, prediction of biotransformation rates is difficult [20].

The importance of biotransformation in drug activity [37] and in assessing human risk of environmental toxicants [38] has led to a growing interest in the metabolic pathways of chemicals in bacteria, fish, mammals and other species [39-44]. Quantitative Structure-Activity Relationships (QSARs) have been developed to predict metabolic rates of drugs as well as environmental pollutants, like pesticides and PAHs. Metabolic rates have also been estimated as the difference between the predicted elimination rate neglecting biotransformation and the observed experimental value [20].

However, up to date no direct comparison has been made between the physicochemical properties of xenobiotics and their metabolites. Yet, such comparisons could shed light on general patterns of metabolism. The objective of the present study was to estimate the difference in lipophilicity, expressed by the octanol-water partition coefficient ( $K_{ow}$ ), between parent compounds and their metabolites for a number of organic pollutants. Parent compounds are usually transformed by enzymes into more polar metabolites to be excreted more rapidly; the present work quantifies this difference. The approach can also be considered as a first indication of increased elimination to be used in exposure and risk assessment if empirical data and refined models are lacking.

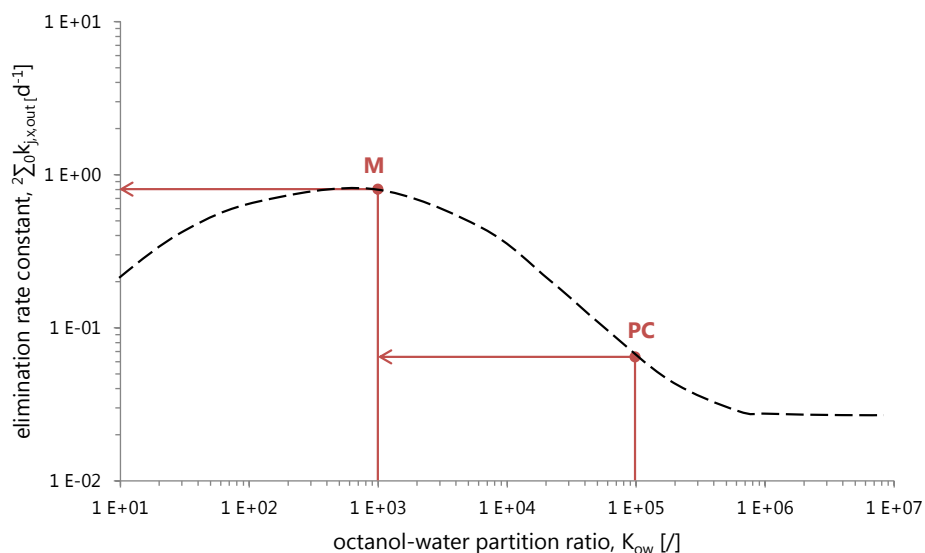
## 2.2 Materials and methods

### 2.2.1 Theory

The octanol-water partition coefficient ( $K_{ow}$ ) is often used in risk assessment to predict intake, accumulation and excretion rates of chemicals [15]. Elimination rate constants for persistent chemicals generally decrease with the  $K_{ow}$  (Figure 2.1) [45, 46]. Biotransformation usually reduces the lipophilicity of the compound, facilitating its excretion via aqueous fluids [47]. If the parent compound is immediately and totally metabolised, it can be assumed that the elimination of the metabolite is similar to that of a persistent compound which is as lipophilic as the metabolite. As an example, Figure 2.1 shows the increase of the elimination rate constant by a factor of 10, from about 0.08 to 0.80, as a

result of the reduction of the  $K_{ow}$  by two orders of magnitude, i.e. from  $10^5$  to  $10^3$ . The dashed line refers to elimination rate constants representing total physical-chemical elimination of persistent compounds, i.e. without biotransformation, in  $10^{-1}$  kg mammals [15].

Figure 2.1. Effect of a  $K_{ow}$  reduction from parent compound (PC) to metabolite (M) on the elimination rate constant. Background graph taken from Hendriks et al. 2001 [15].



### 2.2.2 Data collection

Information on the metabolic pathways of a set of environmental pollutants (parent compounds) was taken from the scientific literature and from two publicly available databases: Hazardous Substances Data Bank (HSDB, <http://toxnet.nlm.nih.gov/>) and Toxin and Toxin Target Database (T3DB, <http://www.t3db.org/>). We built a database including those pollutants that have one main metabolic pathway in mammals and that are oxidised by the enzymes alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH) and cytochrome P450 (P450) [48]. The parent compounds were grouped according to their first “metabolite”, i.e. to the reaction they undergo. We considered the following biotransformation reactions: alcohol oxidation (by ADH), aldehyde oxidation (by ALDH) and the more common types of P450 reactions [49, 50], i.e. hydroxylation, dihydroxylation, epoxidation and heteroatom (N, S) oxygenation. Appendix A provides a scheme with the biotransformation reactions on chemical moieties (Table A1). The parent compounds and the

relative metabolites can be also found in Appendix A (Table A2), together with their Log K<sub>ow</sub> values and literature references.

The octanol-water partition coefficients of parent compounds and metabolites were taken from the ChemSpider database (freely accessible at <http://www.chemspider.com/>). ChemSpider reports the experimental Log K<sub>ow</sub> values (when available in the database), as well as the predicted values calculated by the ACD/logP program [51], without the relative uncertainties. This program has the advantage of accounting for the positional (topological) effect of substituents on a chemical structure [52].

### 2.2.3 Data treatment

The Log-transformed octanol-water partition coefficients of the metabolites, Log K<sub>ow (metabolite)</sub> were related to the parent compounds, Log K<sub>ow (parent)</sub>, according to

$$\text{Log K}_{\text{ow (metabolite)}} = a \cdot \text{Log K}_{\text{ow (parent)}} + b \text{ (Equation 1).}$$

The linear parameters *a* (slope) and *b* (intercept), as well as the statistical standard error (SE), the correlation coefficient (*r*<sup>2</sup>), 95% the confidence interval (95%CI), and the significance level (*p*) were determined. Slopes and intercepts were analysed for significant deviation from *a*=1 and *b*=0, respectively, i.e. from the bisector representing a 1:1 relation between the Log K<sub>ow</sub> values of parent compounds and metabolites.

We developed one regression per enzyme (general regressions) and one per biotransformation reaction. A first set of regressions was built using Log K<sub>ow</sub> values calculated by the ACD/logP program and a second one using experimental Log K<sub>ow</sub> values, when available for at least 5 parent compounds and their relative metabolites. An analysis of covariance (ANCOVA) [53] was performed to compare the regressions with experimental Log K<sub>ow</sub> values with the regression with predicted values. If the *p*<sub>ancova</sub> resulting from the test for homogeneity of regression was lower than 0.05, we considered the two regressions significantly different from each other.

## 2.3 Results

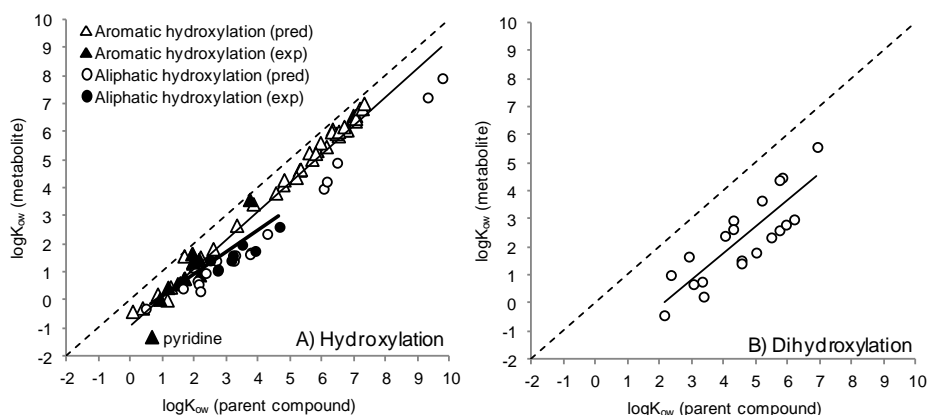
In Figure 2.2, the Log K<sub>ow</sub> of the parent compound is plotted against the Log K<sub>ow</sub> of the metabolite, using calculated (empty symbols, thin lines) and experimental values (full symbols, thick lines). Tables 2.1 and 2.2 provide the regression equations and statistical parameters obtained for all metabolic pathways considered, using calculated and experimental Log K<sub>ow</sub> values, respectively. All regressions were significant at the 0.01 level (*p*<0.01).



The regressions with predicted Log  $K_{ow}$  had high correlation coefficients:  $r^2$  was higher than 0.85, except for dihydroxylation ( $r^2=0.71$ ) and N-hydroxylation ( $r^2=0.63$ ). The slopes were equal to 1 within a 95% CI. The general regression lines (Figure 2.2) gathered around the intercepts  $b=0$  (ADH and ALDH) and  $b=-1$  (CYP), indicating metabolic pathways that do not change the  $K_{ow}$  of substrates and metabolic pathways that lower the  $K_{ow}$  by a factor of 10, respectively. More in detail (Table 2.1), for hydroxylation and epoxidation the intercept was statistically similar to -1 within a 95% CI, while for dihydroxylation and sulfoxidation it was around -2. In contrast, the intercepts were about 0 for N-hydroxylation and for the oxidation of alcohols to aldehydes and to ketones.

Using experimental Log  $K_{ow}$  data, we also set up nine validation regressions (Table 2.2 and thick lines in Figure 2.2). These regressions were significant at the 0.01 level, with explained variance ranging from 70 to 99%. The regressions with experimental and with predicted  $K_{ow}$  values were statistically similar, with the exception of aromatic hydroxylation and the regressions mediated by ADH, which had  $p_{ancova}<0.05$ .

Figure 2.2. Log  $K_{ow}$  values of metabolites versus parent compounds, using predicted (empty dots) or experimental (full dots) Log  $K_{ow}$  values, for the following biotransformation reactions: **a.** hydroxylation; **b.** epoxidation; **c.** dihydroxylation; **d.** sulfoxidation; **e.** N-hydroxylation; **f.** oxidation of alcohols to aldehydes; **g.** oxidation of alcohols to ketones; **h.** oxidation of aldehydes. Dashed lines indicate the 1:1 bisector ( $a=1$  and  $b=0$ ), while solid lines indicate the regressions with predicted (thin lines) or experimental (thick lines) Log  $K_{ow}$  values.



(continues on next page)

Continuation of Figure 2.2.

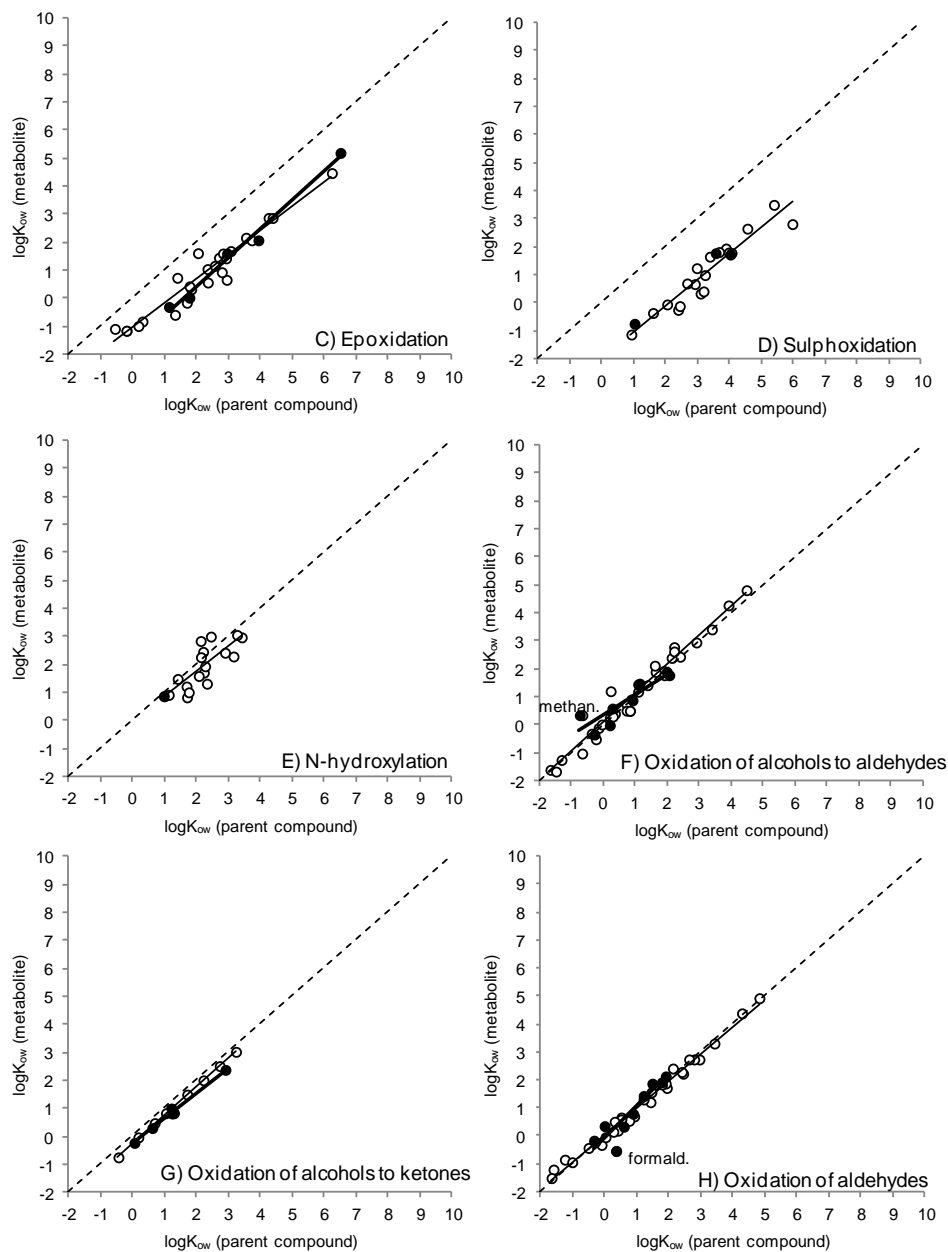


Table 2.1. Characteristics and statistical parameters of metabolite versus parent compound  $\text{Log}(K_{ow})$  regressions with slope  $a$  and intercept  $b$ .  $\text{Log } K_{ow}$  are calculated with the ACD/LogP program.

Metabolic reaction	n	Log $K_{ow}$ range of PCs	a $\pm$ SE	95%CI <sup>a</sup> a	b $\pm$ SE	95%CI <sup>a</sup> b	r <sup>2</sup>	SE	p <sup>b</sup>
P450 enzymes									
General regression	147	-0.56; 9.76	0.97 $\pm$ 0.03	0.90; 1.04	-1.13 $\pm$ 0.14	-1.42; -0.85	0.85	0.86	<0.01
Hydroxylation	65	0.05; 9.76	1.02 $\pm$ 0.03	0.96; 1.08	-0.97 $\pm$ 0.15	-1.27; -0.67	0.95	0.56	<0.01
	47	0.05; 7.31	1.04 $\pm$ 0.01	1.01; 1.06	-0.74 $\pm$ 0.06	-0.86; -0.61	0.99	0.19	<0.01
	18	0.47; 9.76	0.92 $\pm$ 0.03	0.86; 0.97	-1.33 $\pm$ 0.13	-1.61; -1.05	0.99	0.30	<0.01
	20	2.13; 6.91	0.94 $\pm$ 0.14	0.64; 1.23	-1.97 $\pm$ 0.67	-3.38; -0.55	0.71	0.84	<0.01
	25	-0.56; 6.23	0.86 $\pm$ 0.05	0.76; 0.96	-1.04 $\pm$ 0.14	-1.32; -0.76	0.93	0.39	<0.01
	19	0.92; 5.96	0.94 $\pm$ 0.08	0.78; 1.11	-2.02 $\pm$ 0.27	-2.58; -1.45	0.90	0.40	<0.01
	18	0.99; 3.41	0.91 $\pm$ 0.17	0.54; 1.28	-0.07 $\pm$ 0.39	-0.91; 0.76	0.63	0.49	<0.01
ADH									
General regression	43	-1.69; 4.45	1.02 $\pm$ 0.04	0.95; 1.10	0.04 $\pm$ 0.06	-0.09; 0.17	0.95	0.33	<0.01
Oxidation of primary alcohols to aldehydes	33	-1.69; 4.45	1.04 $\pm$ 0.04	0.96; 1.12	0.11 $\pm$ 0.07	-0.02; 0.25	0.96	0.32	<0.01
Oxidation of secondary alcohols to ketones	10	-0.45; 3.23	1.02 $\pm$ 0.02	0.96; 1.07	-0.27 $\pm$ 0.04	-0.37; -0.17	0.99	0.08	<0.01
ALDH									
Oxidation of aldehydes to acids	32	-1.67; 4.82	0.97 $\pm$ 0.03	0.92; 1.03	0.01 $\pm$ 0.05	-0.09; 0.12	0.98	0.24	<0.01

<sup>a</sup> Confidence Interval for the parameter at 95% confidence; <sup>b</sup> p-value of statistical significance testing.

Table 2.2. Characteristics and statistical parameters of metabolite versus parent compound  $\text{Log}(K_{ow})$  regressions with slope a and intercept b.  $\text{Log } K_{ow}$  are experimental values.

Metabolic reaction	n	$\text{Log } K_{ow}$ range of PCs	a $\pm$ SE	95%CI <sup>a</sup> a	b $\pm$ SE	95%CI <sup>a</sup> b	r <sup>2</sup>	SE	p <sup>b</sup>	p <sub>ancova</sub> <sup>c</sup>
P450										
General regression, incl. (N,S) oxygenation	26	0.65; 6.50	0.84 $\pm$ 0.09	0.66; 1.01	-0.86 $\pm$ 0.25	-1.38; -0.35	0.80	0.60	<0.01	0.28
Hydroxylation	17	0.65; 4.66	0.79 $\pm$ 0.13	0.51; 1.07	-0.65 $\pm$ 0.35	-1.40; 0.09	0.71	0.59	<0.01	0.07
	10	0.65; 3.72	1.40 $\pm$ 0.16	1.03; 1.76	-1.52 $\pm$ 0.31	-2.23; -0.80	0.91	0.41	<0.01	<0.01
	9	0.90; 3.72	1.23 $\pm$ 0.13	0.92; 1.55	-1.13 $\pm$ 0.27	-1.77; -0.48	0.92	0.31	<0.01	0.04
	7	2.49; 4.66	0.61 $\pm$ 0.13	0.28; 0.94	-0.35 $\pm$ 0.44	-1.48; 0.77	0.82	0.23	<0.01	0.06
Epoxidation	5	1.13; 6.50	1.03 $\pm$ 0.07	0.82; 1.24	-1.64 $\pm$ 0.25	-2.42; -0.86	0.99	0.28	<0.01	0.10
ADH										
General regression	14	-0.77; 2.90	0.74 $\pm$ 0.10	0.52; 0.97	0.17 $\pm$ 0.13	-0.13; 0.46	0.82	0.37	<0.01	0.01
Oxidation of primary alcohols to aldehydes without methanol	8	-0.77; 2.03	0.74 $\pm$ 0.15	0.38; 1.09	0.34 $\pm$ 0.17	-0.07; 0.75	0.81	0.39	<0.01	0.02
	7	-0.31; 2.03	0.94 $\pm$ 0.12	0.64; 1.24	0.08 $\pm$ 0.14	-0.28; 0.45	0.93	0.29	<0.01	0.31
Oxidation of secondary alcohols to ketones	6	0.05; 2.90	0.91 $\pm$ 0.04	0.80; 1.02	-0.26 $\pm$ 0.06	-0.43; -0.10	0.99	0.09	<0.01	0.04
ALDH										
Oxidation of aldehydes to acids	9	-0.34; 1.90	1.14 $\pm$ 0.19	0.70; 1.58	-0.09 $\pm$ 0.21	-0.59; 0.41	0.84	0.41	<0.01	0.21
without formaldehyde	8	-0.34; 1.90	1.03 $\pm$ 0.11	0.76; 1.30	0.11 $\pm$ 0.13	-0.21; 0.44	0.94	0.24	<0.01	0.62

<sup>a</sup> Confidence Interval for the parameter at 95% confidence; <sup>b</sup> p-value of statistical significance testing; <sup>c</sup> p-value of statistical homogeneity of regression testing.

## 2.4 Discussion

### 2.4.1 Calculation methodology

In this study, we related the  $\text{Log } K_{ow}$  of parent compounds to the  $\text{Log } K_{ow}$  of their first metabolites in mammals, dividing the data according to the metabolic pathway. We also built general regressions merging data per enzyme group (CYP, ADH, ALDH).

All regressions developed with predicted  $\text{Log } K_{ow}$  values were robust and statistically significant and had slopes containing the value of 1 in their 95% confidence intervals (Table 2.1). The dispersion of the data in Figure 2.2 (empty symbols, thin lines) was generally similar both at low and high  $K_{ow}$ , indicating that the total lipophilicity depends on electronic interactions among substituents of the chemical structure. Errors and uncertainties affecting the calculated values of  $K_{ow}$  were not provided by the Chemspider database. Nevertheless, as the same error affects both parent compounds and metabolites, the pattern still remains consistent.

The interpretation of the results is closely related to the method used to calculate the  $K_{ow}$ . Since the octanol-water partition coefficient has long been known as an “additive-constitutive” property [51], the ACD/logP software uses the basic approach of “group contribution”, which is valid among different chemical classes and in a large range of  $\text{Log } K_{ow}$  values. If a metabolic process effectively “removes” a group of atoms and “inserts” a different one, the overall lipophilicity change will depend only on the difference between the contribution of both group. For this assumption, each regression is expected to have a slope of exactly one, as the difference is independent of the total lipophilicity of the molecule. In other words, Equation 1 can be considered in terms of a Hammett equation:  $\text{Log}(K_{ow(\text{metabolite})}/K_{ow(\text{parent})})=b$ . In this equation  $K_{ow}$  coefficients are equilibrium constants which can be related to free energies of solvation by simple thermodynamical laws. Thus, the difference between  $\text{Log } K_{ow}$  becomes the difference between free energies of solvation of the metabolite and parent compound. The intercept “b” is negative when  $K_{ow(\text{metabolite})} < K_{ow(\text{parent})}$  and positive when  $K_{ow(\text{metabolite})} > K_{ow(\text{parent})}$ . In Hammett terms (“total electronic effect”) this means that the insertion of an oxygen atom or link has a favouring or disfavouring electronic effect on the solvation by water. Usually, this insertion favours the water solubility for several reasons: raised molecular volume, raised H-bond basicity, raised polarizability, etc. Thus, the intercept “b” is expected to be negative for the oxidation reactions considered in our study.

We set up 9 validation regressions using experimental  $\text{Log } K_{ow}$  values and analysed their similarity to the regressions with predicted  $\text{Log } K_{ow}$ . The  $p_{\text{ancova}}$

resulting from the analysis of covariance (Table 2.2) confirmed the homogeneity between the two types of regressions, with the exceptions of aromatic hydroxylation and the regressions for ADH, with  $p_{\text{ancova}} < 0.05$ . Figures 2.2a and 2.2f show deviations for two data points: pyridine and methanol (experimental Log  $K_{\text{ow}}$  values), undergoing aromatic hydroxylation and alcohol oxidation, respectively. It is interesting to note that formaldehyde presented a deviation in the regression for ALDH compounds with experimental Log  $K_{\text{ow}}$  data (Figure 2.2h). Formaldehyde ( $\text{CH}_2\text{O}$ ) and methanol ( $\text{CH}_3\text{OH}$ ) are the simplest aldehyde and the simplest alcohol, respectively. Thus, these molecules may not adhere to general trends because of their small size. In order to test the sensitivity, regressions were developed removing pyridine, methanol and formaldehyde from their respective datasets with experimental Log  $K_{\text{ow}}$ . The results are reported in Table 2.2: the fit was improved, as well as the homogeneity of the regressions (higher  $p_{\text{ancova}}$ ).

#### 2.4.2 Intercepts

The regression lines reflect an increase (intercept  $> 0$ ) or decrease (intercept  $< 0$ ) of the lipophilicity after biotransformation. The oxidation reactions of alcohols and aldehydes did not lead to a significant lipophilicity change, having intercepts of about zero. While this may be at odds with the high metabolic rates usually noted for alcohols [48], one has to keep in mind that this hydrophobicity trend allows the reverse reduction of aldehydes to alcohols driven by the alcohol dehydrogenase [54]. Furthermore, the majority of acids deprotonate at cytosolic pH, the ionic form being more water-soluble, thus more easily excretable.

The decrease in lipophilicity differed for the single reactions mediated by CYP enzymes. Hydroxylation and epoxidation reduced the lipophilicity by one order of magnitude ( $b = -0.97$  and  $b = -1.04$ , respectively). Dihydroxylation and sulfoxidation reduced the  $K_{\text{ow}}$  by two orders of magnitude ( $b = -1.97$  and  $-2.02$ , respectively). The two orders of magnitude difference for sulfoxidation was confirmed by a similar study on the oxidation of alkyl sulphides [55]. Experimental Log  $K_{\text{ow}}$  values of eight phenyl and biphenyl alkyl amines (tertiary) were a linear function of their N-oxidised metabolites in a neutral form, with  $r^2 = 0.93$  and  $p < 0.01$  [56]. Caron et al. concluded that the neutral N-oxides had a Log  $K_{\text{ow}}$  value lower than that of the parent amine by a factor ranging from 2.61 and 2.77. This decrease is higher than those observed with our correlations, due to the differences in chemical structure with respect to the chemicals in this study's dataset. We analysed the N-oxygenation of primary and secondary amines to hydroxylamines, which is the only reaction mediated by CYP enzymes that cause no change in Log  $K_{\text{ow}}$ , with the intercept close to zero. Overall, Log  $K_{\text{ow}}$  was shown to be reduced by one unit for

chemicals that are typically metabolised by CYP, the intercept being -1.13. The biotransformation reactions considered in the present study are the more common reactions mediated by CYP enzymes.

The excretion of stable compounds decreases with hydrophobicity [15]. Vice versa, a reduction of the  $K_{ow}$  by biotransformation will thus enhance elimination to an extent that may be anticipated by the same relationship (Figure 2.1). Obviously, empirical confirmation by future studies is needed. As metabolism rates are hard to anticipate with existing methods, we feel that the present paper provides the first necessary step in an alternative approach [20].

## 2.5 Conclusions

Comparisons of lipophilicity and preliminary discussions on their significance play a key role in understanding the natural logic of metabolism. The present study shows that the  $\text{Log } K_{ow}$  is reduced by a factor that varies between 0 and -2, depending on the metabolic pathway. The magnitude of the reduction can be anticipated by analysing the way the  $K_{ow}$  is calculated. Knowing the magnitude of the reduction is a first necessary step in an alternative approach to estimating biotransformation rates.

## Appendix

Appendix A provides a scheme with the biotransformation reactions on chemical moieties (Table A1). The parent compounds and the relative metabolites can be found in Table A2 of Appendix A, together with their  $\text{Log } K_{ow}$  values and literature references.

## Acknowledgments

Financial support by the European Union through the Environmental ChemOinformatics (ECO) project (FP7-PEOPLE-ITN-2008, n. 238701) is gratefully acknowledged.





## Chapter 3

# **Compound lipophilicity as a descriptor to predict binding affinity ( $1/K_m$ ) in mammals**

Alessandra Pirovano

Mark A.J. Huijbregts

Ad M.J. Ragas

A. Jan Hendriks

*Environmental Science & Technology* (2012), 46(9), 5168-5174

### 3.1 Introduction

The EU REACH (Registration, Evaluation, Authorization and restriction of CHemicals) legislation [2] requires the risk assessment of thousands of chemicals to evaluate the potential adverse effects that exposure to chemicals may have on human health and the environment. Due to financial, practical and ethical constraints, not all compounds can be tested on all species to be protected. Thus, models are needed to predict fate and effects of new and existing chemicals [7].

The accumulation of xenobiotics in organisms is a key factor in the risk assessment of chemicals. In bioaccumulation models, biotransformation is one of the processes decreasing the concentration of chemicals in an organism, together with elimination through physicochemical processes, e.g. excretion via water, egestion via faeces and growth dilution [15]. Parent compounds can be transformed via enzymatic reactions to metabolites, which are usually more polar and can thus be excreted more easily. The enzymatic action of metabolism involves two processes. Firstly, the chemical needs to reach the enzyme and bind to it; secondly, a catalytic reaction has to take place. The binding of the chemical and its successive catalysis are described by two enzymatic parameters: the Michaelis constant ( $K_m$ ) and the maximum rate of the reaction ( $V_{max}$ ), respectively [25]. The  $K_m$  value is the substrate concentration at half the maximum rate, i.e. at  $V_{max}/2$ , and is independent of the enzyme concentration [1]. The inverse of the Michaelis constant, i.e.  $1/K_m$ , reflects the affinity of the enzyme for its substrate: a low  $K_m$  (or high  $1/K_m$ ) corresponds to high binding affinity.

Measured  $K_m$  and  $V_{max}$  data are lacking for many chemicals and species. Models based on experimental data can be used to predict the biological activity of a broader range of related chemicals. So far, QSARs have been developed to explore the relationships between the enzymatic constants ( $K_m$  and  $V_{max}$ ) and substrate characteristics with regard to drugs oxidised by the microsomal cytochrome P450 (CYP) [26, 27]. The affinity, represented by  $1/K_m$ , was shown to be mainly related to the lipophilicity of the compound (see reviews [25, 28]), although other factors might also be important, such as ionic interactions and hydrogen bonding properties [30]. However, these models focussed on single CYP isoenzymes and small datasets, mainly drugs. We investigated the relationship between affinity and lipophilicity extending the analysis to a broader set of chemicals. CYP is the major (and thus the most studied) enzyme group in terms of catalytic versatility and the large number of xenobiotics it detoxifies or activates [13]. Nevertheless, the contribution of other enzymes to the oxidative metabolism of xenobiotics is significant as well [32]. Despite their importance, QSARs for non-CYP enzymes have not been developed. We

hypothesised that the lipophilicity-binding regressions found for small datasets of CYP substrates could be extended to non-CYP enzymes.

The aim of this study was therefore to estimate the relationships between  $K_m$  and lipophilicity, expressed by the octanol-water partitioning coefficient ( $K_{ow}$ ), in mammals. Regressions were developed for oxidations catalysed by alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), flavin-containing monooxygenase (FMO) and CYP enzymes, in order to find generic patterns of metabolism across enzymes.

## 3.2 Methods

### 3.2.1 Data selection

Michaelis constants ( $K_m$ ) were collected for alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH) and flavin-containing monooxygenase (FMO). For ADH and ALDH, data were taken from the BRENDA enzyme database (BRAunschweig ENzyme DAtabase, <http://www.brenda-enzymes.org>) [57].  $K_m$  values for FMO were taken from a review [58] and references contained therein. We also collected  $K_m$  values for cytochrome P450 (CYP) from reviews [26, 59, 60]. All data extracted from the BRENDA database and the reviews were checked in the original papers. We assumed that  $K_m$  data were of adequate quality as taken from peer reviewed articles.

Michaelis constants ( $K_m$ , reported in  $\mu\text{M}$ ) were combined into four databases, one for each enzyme family. Inclusion criteria were as follows:  $K_m$  measured for mammals in *in vitro* assays of purified, non-recombinant, hepatic enzymes. For every  $K_m$  value, we recorded the species and the enzyme for which it was measured, and the experimental conditions such as pH and temperature.

SMILES (Simplified Molecular Input Line Entry System) strings [61] and CAS (Chemical Abstract Service) numbers were obtained for each compound from the ChemSpider website (<http://www.chemspider.com/>). The octanol-water partitioning coefficients ( $K_{ow}$ ) were taken from the KOWWIN<sup>TM</sup> v 1.67, a program of EPI Suite<sup>TM</sup> available at the website of US EPA (Environmental Protection Agency <http://www.epa.gov>). Experimental  $K_{ow}$  values, when available, were preferred over estimated ones. As the datasets included a number of compounds that would be ionised at physiological pH (7.4), we obtained Log  $D_{7.4}$  values from ChemSpider, which are calculated using the software ACD Laboratories LogD (Advanced Chemistry Development ACD/Laboratories Research, Toronto, Canada). The distribution coefficient  $D_{7.4}$  represents the partitioning coefficient corrected for ionisation of the chemical at pH 7.4.

Each compound was assigned to relevant chemical classes using the ECOSAR<sup>TM</sup> program v 1.0 present in EPI Suite<sup>TM</sup>. ECOSAR recognises the presence of specific functional groups denoting the compound. If the functional group is detected then the compound is allocated into the respective class(es) [62].

The  $K_m$  data collected can be found in Appendix B (Table B1), with the references to the original papers.

### 3.2.2 Data treatment

For each enzyme family, data were grouped per species (i.e. human, horse, rat, mouse, pig and rabbit) and isoenzymes. The isoenzymes are any of the several forms of an enzyme, all of which catalyse the same reaction but are characterised by varying properties (e.g. electrophoresis, chromatography, kinetics criteria, chemical structure, etc). Regressions were developed for each combination of a species and isoenzyme (specific regressions). In addition, all species and isoenzymes were merged into one regression per enzyme family (general regression).

Each substrate was characterised by a single value in order to prevent bias due to the overrepresentation of  $K_m$  values of substrates which were measured either in different species and/or isoenzymes, or more than one time in the same combination of species and isoenzyme. For this purpose, if multiple values were available for one substrate, we calculated the geometric mean of the experimental  $K_m$  values, as well as the geometric standard deviation.

### 3.2.3 Data analysis

Linear regression analysis was performed using the Ordinary Least Squares (OLS) method. Among all datasets built with the different combinations of species/isoenzymes, we included in the analysis only those containing at least 6 compounds. For each dataset, the QSAR equations were developed in the form:

$$\text{Log}(1/K_m) = a \cdot \text{Log } K_{ow} + b \quad (\text{Eq. 3.1})$$

We reported the slope (a) and the intercept (b) with their standard errors. The quality of the regression was characterised by the number of compounds used in the model (n), coefficient of determination ( $r^2$ ), standard error for the estimated parameter  $\text{Log}(1/K_m)$  (SE) and the *p-value* from the F-test (p). We also calculated the 95% Confidence Interval (95%CI) for slopes and intercepts. In order to explore the influence of ionisation in enzyme binding, we also developed the general regressions for the four enzyme families using  $\text{Log } D_{7.4}$  values instead of  $\text{Log } K_{ow}$ .

An analysis of covariance (ANCOVA) was performed to compare every specific regression with the general regression, within an enzyme group. If the

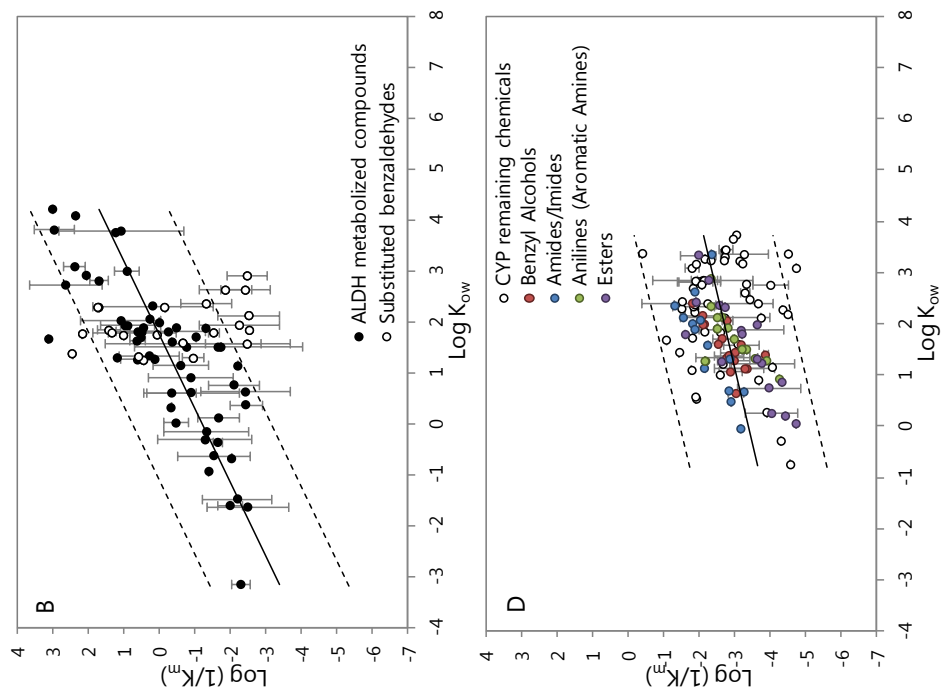
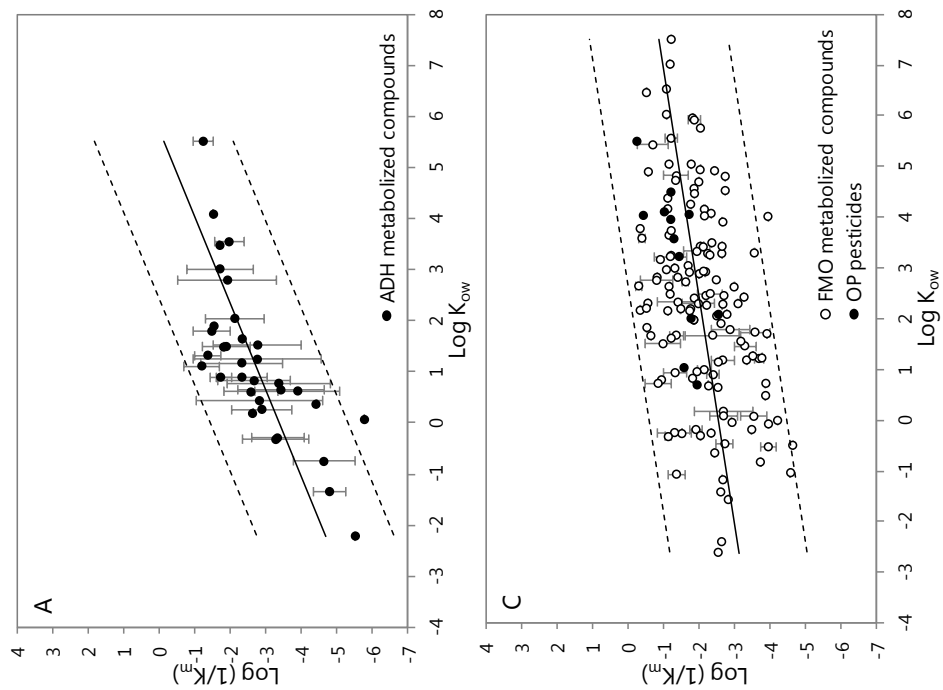
resulting  $p_{\text{ancova}}$  was lower than 0.05, we considered that the two regressions significantly differ from each other.

In addition, separate regressions were developed for specific groupings of compounds metabolised by FMO and CYP for which we expected a similar behaviour. The FMO database contains several chemicals that are used as pesticides and are biologically highly active: 12 organophosphorous (OP), 4 carbamate (CM) and 5 dithiocarbamate (DTC) compounds. A list of these compounds is reported in Appendix C (Table C1), together with their original ECOSAR classes and their general structure. The ECOSAR software does not separately categorise reactive chemicals such as OPs and CMs [63]. Therefore, we manually classified them and made a separate regression for OP pesticides, the only group with more than 6 compounds. For CYP, which has a wide substrate specificity, regressions were developed for single ECOSAR classes, or for combinations of similar classes: Anilines (Aromatic Amines), Benzyl Alcohols, Esters and Amides/Imides. The compounds that did not belong to these well-defined classes were combined in a group called 'remaining chemicals'. The vast majority of the chemicals in this group belong to the ECOSAR class Neutral Organics. The ECOSAR software defines Neutral Organics as compounds that are generally solvents, non-ionisable and non-reactive [63], thus including diverse chemicals.

### 3.3 Results

All regressions made for each combination of isoenzyme and/or species are reported in Tables 3.1-3.4, corresponding to ADH, ALDH, FMO and CYP, respectively. From here on, the equations are specified with their names, which describe the enzyme family and the isoenzyme, indicated by its number and/or the species, indicated by its first 3 letters. Appendix C (Tables C2-C5) provides a more complete overview of the regressions, including the 95% Confidence Interval (95%CI) for slopes and intercepts, as well as the Log ( $1/K_m$ ) and Log  $K_{ow}$  ranges.

Figure 3.1 (next page). Relationships between Log ( $1/K_m$ ) and Log  $K_{ow}$  in mammals for compounds metabolised by: A) ADH; B) ALDH; C) FMO; D) CYP. Regressions (solid lines) and 95% confidence intervals (dashed lines). Laboratory measurements (dots): Log transformed geometrical mean of  $1/K_m$  [ $\mu\text{M}^{-1}$ ] for each compound, with the geometric standard deviation (vertical bar).



### 3.3.1 ADH

We developed 7 equations for ADH, which are reported in Table 3.1. The slope of the general regression ADHgen (Figure 3.1A) was 0.6, and the observed  $K_m$  data were between  $10$  and  $10^6$   $\mu\text{M}$ . The specific regressions had a systematically lower explained variance compared to ADHgen ( $r^2=0.56$ ), except for ADH3\_rat which had an  $r^2$  of 0.77. With  $p_{\text{ancova}} < 0.05$ , the 2 regressions for ADH3 were statistically different from the general one; in particular, the intercepts were smaller. ADH dataset contained a large number of compounds classified as Neutral Organics (17 on a total of 33). They were mainly linear alcohols, while two compounds were classified as Benzyl Alcohols.

Table 3.1. Relationships between Log  $K_{ow}$  and Log  $(1/K_m)$  for ADH. The  $K_m$  values were expressed as  $\mu\text{M}$ .

Name	Slope( $\pm$ SE)	Intercept( $\pm$ SE)	n	$r^2$	SE	$p^a$	$p_{\text{ancova}}^b$
Regression made merging all species (mammals) and all isoenzymes							
ADHgen	0.59( $\pm$ 0.09)	-3.36( $\pm$ 0.18)	34	0.56	0.82	<0.01	/
Regressions for the separate species and the separate isoenzymes							
ADH1_hor	0.40( $\pm$ 0.11)	-3.08( $\pm$ 0.24)	20	0.45	0.72	<0.01	0.96
ADH1_hum	0.58( $\pm$ 0.12)	-3.01( $\pm$ 0.23)	24	0.50	0.85	<0.01	0.13
ADH2_hum	0.67( $\pm$ 0.19)	-3.58( $\pm$ 0.41)	18	0.43	1.43	<0.01	0.70
ADH3_hum	0.54( $\pm$ 0.25)	-4.38( $\pm$ 0.58)	7	0.48	0.72	<u>0.09</u>	<u>&lt;0.01</u>
ADH1_rat	0.62( $\pm$ 0.19)	-3.11( $\pm$ 0.32)	13	0.50	0.84	0.01	0.28
ADH3_rat	1.18( $\pm$ 0.32)	-6.57( $\pm$ 0.75)	6	0.77	0.82	0.02	<u>&lt;0.01</u>
Regression made merging all species and all isoenzymes, using Log $D_{7.4}$ values							
ADHgen ionis	0.60( $\pm$ 0.10)	-3.30( $\pm$ 0.18)	34	0.52	0.85	<0.01	/

<sup>a</sup> The underlined value indicates non significant regression ( $p > 0.05$ ); <sup>b</sup> the underlined values indicate regressions significantly different from ADHgen ( $p_{\text{ancova}} < 0.05$ ).

### 3.3.2 ALDH

We initially built 9 QSARs for ALDH, which are reported in Table 3.2. The general equation ALDHgen (Figure 3.1b) had a slope of 0.7, and the observed  $K_m$  data were between  $10^{-3}$  and  $10^3$   $\mu\text{M}$ . Among the specific regressions, the 3 equations for rat had  $r^2$  values lower than for human and horse ( $r^2$  between 0.4 and 0.8). Compared to ALDHgen, the 3 equations for rat had  $p_{\text{ancova}} < 0.05$ . For ALDHgen, 11 out of the total 77 compounds had observed  $K_m$  values that were 2 orders of magnitude larger or smaller than expected from the regression.

Nine of these outliers were substituted benzaldehydes. The ALDH dataset contained 22 substituted benzaldehydes, which are represented by white dots in Figure 3.1B and listed in Appendix C (Table C6), together with their general structures.

We developed 3 additional general regressions leaving out the possibly influential data: I) substituted benzaldehydes; II) rat data; III) rat data as well as substituted benzaldehydes. The 3 additional regressions (Table 3.2) had a slope of 0.8 and  $r^2$  values larger than ALDHgen ( $r^2=0.33$ ). The exclusion of the substituted benzaldehydes significantly improved the correlation: the explained variance was increased to 63%, and SE was reduced from 1.33 to 0.96. Similar statistic parameters were obtained when both rat data and substituted benzaldehydes were removed from the dataset. In order to discern the contribution of rat data to the weak correlations found for ALDH, we developed two more regressions: 1) including only rat data for ALDH metabolised compounds; 2) including only rat data and excluding substituted benzaldehydes. The results are reported in Appendix C (Table C7, Figure C1). No robust correlation was found between Log  $K_{ow}$  and Log ( $1/K_m$ ) in rat, with explained variance of 6% and a slope of 0.16. The correlation was improved by the exclusion of substituted benzaldehydes, although it was still weak ( $r^2=0.28$ ).

Table 3.2. Relationships between Log  $K_{ow}$  and Log ( $1/K_m$ ) for ALDH, together with 3 additional general regressions leaving out the possibly influential data: I) substituted benzaldehydes; II) rat data; III) rat data as well as substituted benzaldehydes. The  $K_m$  values were expressed as  $\mu M$ .

Name	Slope( $\pm$ SE)	Intercept( $\pm$ SE)	n	$r^2$	SE	$p^a$	$p_{\text{ancova}}^b$
Regression made merging all species (mammals) and all isoenzymes							
ALDHgen	0.69( $\pm$ 0.11)	-1.18( $\pm$ 0.22)	77	0.33	1.33	<0.01	/
Regressions for the separate species (mammals) and the separate isoenzymes							
ALDH1_hor	0.99( $\pm$ 0.30)	-1.31( $\pm$ 0.38)	10	0.57	1.00	0.01	0.84
ALDH2_hor	0.73( $\pm$ 0.35)	-0.43( $\pm$ 0.43)	9	0.39	1.13	<u>0.07</u>	0.10
ALDH1_hum	0.82( $\pm$ 0.08)	-0.99( $\pm$ 0.17)	28	0.80	0.73	<0.01	0.19
ALDH2_hum	0.86( $\pm$ 0.13)	-0.73( $\pm$ 0.27)	57	0.42	1.17	<0.01	<u>&lt;0.01</u>
ALDH3_hum	0.54( $\pm$ 0.17)	-1.18( $\pm$ 0.21)	12	0.51	0.74	0.01	0.95
ALDH1_rat	0.18( $\pm$ 0.10)	-1.33( $\pm$ 0.17)	32	0.10	0.73	<u>0.08</u>	<u>&lt;0.01</u>
ALDH2_rat	0.10( $\pm$ 0.17)	-2.34( $\pm$ 0.26)	22	0.02	1.00	<u>0.55</u>	<u>&lt;0.01</u>
ALDH3_rat	0.56( $\pm$ 0.33)	-3.80( $\pm$ 0.74)	8	0.32	0.45	<u>0.14</u>	<u>&lt;0.01</u>



### Continuation of Table 3.2

Additional general regressions excluding possibly influential data: I) substituted benzaldehydes; II) rat data; III) rat data and substituted benzaldehydes s							
I	0.81(±0.09)	-1.15(±0.17)	55	0.63	0.96	<0.01	/
II	0.83(±0.10)	-0.84(±0.20)	63	0.53	1.05	<0.01	/
III	0.83(±0.09)	-0.92(±0.19)	50	0.63	0.96	<0.01	/
Regression made merging all species and all isoenzymes, using Log D <sub>7.4</sub> values							
ALDHgen ionis	0.61(±0.12)	-1.00(±0.23)	77	0.26	1.4	<0.01	/

The underlined values indicate: <sup>a</sup>non significant regressions ( $p > 0.05$ )  
<sup>b</sup>regressions significantly different from ALDHgen ( $p_{\text{ancova}} < 0.05$ ).

#### 3.3.3 FMO

In most of the experiments in which FMO activity was measured, the isoenzyme investigated was not reported. Thus, it was possible to group the data by species (i.e. mouse and pig) only. For all 3 groupings (Table 3.3), no robust correlations were found between Log  $K_{ow}$  and Log ( $1/K_m$ ), with  $r^2$  values around 0.20. The general equation FMOgen (Figure 3.1C) had a slope of 0.2, and the observed  $K_m$  data were between 1 and  $10^5$   $\mu$ M. With 54% explained variance, the Log  $K_{ow}$  correlated well with the affinity of OP pesticides (represented by black dots in Figure 3.1c), albeit with a shallow slope of 0.3.

Table 3.3. Relationships between Log  $K_{ow}$  and Log ( $1/K_m$ ) for FMO, together with an additional regression developed including organophosphorous (OP) pesticides only. The  $K_m$  values were expressed as  $\mu$ M.

Name	Slope(±SE)	Intercept(±SE)	n	$r^2$	SE	p	$p_{\text{ancova}}$
Regression made merging all species (mammals) and all isoenzymes							
FMOgen	0.22(±0.04)	-2.52(±0.11)	149	0.20	0.88	<0.01	/
Regressions for the separate species							
FMO_mou	0.21(±0.06)	-2.24(±0.16)	45	0.23	0.80	<0.01	0.08
FMO_pig	0.21(±0.04)	-2.48(±0.12)	144	0.18	0.90	<0.01	0.80
Regression for OP pesticides, merging all species and all isoenzymes							
	0.32(±0.09)	-2.34(±0.33)	12	0.54	0.45	0.01	/
Regression made merging all species and all isoenzymes, using Log D <sub>7.4</sub> values							
FMOgen ionis	0.29(±0.04)	-2.43(±0.09)	148 <sup>a</sup>	0.31	0.82	<0.01	/

<sup>a</sup> The Log D<sub>7.4</sub> value of one compound (2-aminoazulene) was not available.

### 3.3.4 CYP

For CYP, we first built 5 QSARs using all data (Table 3.4). The general equation CYPgen had a slope of 0.3 (Figure 3.1D); the observed  $K_m$  data were between 1 and  $10^5$   $\mu\text{M}$ . Among the separate regressions for the ECOSAR classes, poor correlation was found for the group of diverse chemicals, 'remaining chemicals', with  $r^2 < 0.1$  and a slope of 0.2. Good correlations were found for the specific chemical classes, all significant at the 0.01 level and with  $r^2$  values ranging from 0.37 and 0.70. These regressions had slopes between 0.5 and 0.8.

Table 3.4. Relationships between Log  $K_{ow}$  and Log ( $1/K_m$ ) for CYP, together with 5 additional general regressions for separate ECOSAR classes: I) Anilines (Aromatic Amines); II) Benzyl Alcohols; III) Esters; IV) Amides/Imides; V) 'remaining chemicals'. The  $K_m$  values were expressed as  $\mu\text{M}$ .

Name	Slope( $\pm$ SE)	Intercept( $\pm$ SE)	n	$r^2$	SE	$p^a$	$p_{\text{ancova}}$
Regression made merging all species (mammals) and all isoenzymes							
CYPgen	0.34( $\pm$ 0.08)	-3.38( $\pm$ 0.17)	121	0.13	0.82	<0.01	/
Regressions made for the separate species and the separate isoenzymes							
CYP1A1_rat	0.52( $\pm$ 0.17)	-3.63( $\pm$ 0.32)	23	0.30	0.54	0.01	0.75
CYP2B1_rat	0.08( $\pm$ 0.21)	-2.55( $\pm$ 0.48)	39	0.00	1.02	<u>0.70</u>	0.09
CYP2B4_rab	0.24( $\pm$ 0.12)	-3.39( $\pm$ 0.27)	47	0.08	0.76	<u>0.05</u>	0.12
CYP2E1_rab	0.78( $\pm$ 0.10)	-4.00( $\pm$ 0.16)	36	0.65	0.51	<0.01	0.94
I. Regression for Anilines (Aromatic Amines), merging all species and all isoenzymes							
	0.77( $\pm$ 0.26)	-4.19( $\pm$ 0.46)	17	0.37	0.51	0.01	/
II. Regression for Benzyl Alcohols, merging all species and all isoenzymes							
	0.84( $\pm$ 0.20)	-4.03( $\pm$ 0.32)	17	0.54	0.37	<0.01	/
III. Regression for Esters, merging all species and all isoenzymes							
	0.84( $\pm$ 0.14)	-4.48( $\pm$ 0.26)	17	0.70	0.54	<0.01	/
IV. Regression for Amides/Imides, merging all species and all isoenzymes							
	0.48( $\pm$ 0.13)	-3.03( $\pm$ 0.23)	14	0.54	0.43	0.01	/
V. Regression for the remaining chemicals, merging all species and all isoenzymes							
	0.16( $\pm$ 0.13)	-3.02( $\pm$ 0.33)	56	0.03	0.99	<u>0.22</u>	/
Regression made merging all species and all isoenzymes, using Log $D_{7.4}$ values							
CYPgen ionis	0.25( $\pm$ 0.07)	-3.20( $\pm$ 0.15)	121	0.10	0.83	<0.01	/

<sup>a</sup> The underlined values indicate non significant regressions ( $p > 0.05$ ).

### 3.3.4 Ionisation

The general regressions developed for the four enzyme families using Log  $D_{7.4}$  values are reported in the last row of Tables 1-4, as well as in details in Appendix C (Table C8 and Figure C2). The 54% of the compounds in FMO dataset had a dissociated fraction larger than 0.05 at pH 7.4; for the other enzyme families this percentage was 9% or lower. The correction for ionisation improved the results only for FMO, although the correlation was still weak with a slope of 0.3 and  $r^2 = 0.31$ .

## 3.4 Discussion

### 3.4.1 Regressions

The QSAR models presented in this paper were developed for a well-defined endpoint ( $K_m$ ), using an unambiguous algorithm that can be mechanistically interpreted, as recommended by OECD guidelines [64]. The relationship between  $K_{ow}$  and  $1/K_m$  can be understood from partitioning theory. If weak interactions are dominant, the partitioning of organic chemicals over various phases is governed by hydrophobicity and polarity [65]. The lipophilicity parameter Log  $K_{ow}$  combines these two properties [66]. A linear correlation was found between Log  $K_{ow}$  and enzyme binding affinity, expressed as Log ( $1/K_m$ ), similar to the lipophilicity relationships noted for affinity to proteins [65]. The binding affinity increased with the compound  $K_{ow}$  for 4 oxidising enzymes tested *in vitro* in mammals (Tables 3.1-3.4), i.e. the more lipophilic the substrate, the higher its affinity for the enzymes. However, a substantial number of correlations were weak and several were not statistically significant. In such cases, binding affinity may be mainly controlled by other interactions, e.g. of steric, covalent, or ionic nature. Therefore, the inclusion of descriptors related to these components may improve the QSARs.

When available, we used experimental  $K_{ow}$  data, otherwise the predicted ones [26]. The Michaelis constants ( $K_m$ ) were sourced from the open literature, so they come from different laboratories, often employing different protocols (e.g. conditions of pH and temperature) [67]. Consequently, the input data are subject to variation, implying uncertainty in the regressions.

The datasets consisted of specific chemicals; in fact, the experimental  $K_m$  data were taken from tests with compounds considered substrates of the enzymes. The applicability domains of the models are defined by the range (min and max) of Log  $K_{ow}$  values of the compounds used to build the model, which are reported in Tables C2-C5 in Appendix C. Therefore, when using a regression for predicting the  $K_m$  value of a new compound, it is important to know if the chemical is a putative substrate for the enzyme and if its Log  $K_{ow}$  value lies

within the range established by the dataset. Furthermore, it is also recommended to check if the chemical belongs to one of the ECOSAR classes present in the dataset.

We developed 24 QSARs, grouping the data according to 2 criteria: merging all species and all isoenzymes (4 general regressions, one for each enzyme group), and separating each combination of a species and isoenzyme. In most cases, the 4 general QSARs did not differ statistically from the specific ones: apparently, the patterns are generally applicable to different isoenzymes and species. The most remarkable exceptions were the equation for ADH3 and the 3 equations for ALDH in rat. In a previous study on ADH kinetics [68], class 1, 2 and 3 isoenzymes were shown to have common characteristics, such as substrate binding enhancement with increasing compound lipophilicity. Nevertheless, ADH3 is unique among the members of the ADH family, having kinetic properties identical to the glutathione-dependent formaldehyde dehydrogenase [69]. Regarding the regressions for ALDH in rat,  $\text{Log } K_{ow}$  and  $\text{Log } (1/K_m)$  were not strongly correlated. This may explain the difference with the general regression, built using also data from human and horse for which better correlations were found.

We took into account the substrate's dissociation at physiological pH (7.4) by using  $\text{Log } D_{7.4}$  as descriptor, which represents the lipophilicity corrected for ionisation of the chemical. The influence of ionisation to binding affinity was relevant only for compounds metabolised by FMO, for which the correlation with binding affinity increased, though slightly ( $r^2 = 0.31$  and slope = 0.3). Therefore, the inclusion of  $\text{Log } D_{7.4}$  did not contribute to improve the results significantly.

#### *3.4.2 Additional regressions*

We developed 9 additional QSARs including or excluding specific data. For ALDH, the general regression improved when rat data were excluded. In addition, it was found that the binding to ALDH of substituted benzaldehydes was not well described by  $\text{Log } K_{ow}$ . These compounds had similar  $\text{Log } K_{ow}$  values, ranging from 1.22 to 2.88, while their  $\text{Log } (1/K_m)$  values covered 5 orders of magnitude, between -2.51 and 2.49. In the work of Klyosov [70], the kinetics of ALDH towards various aldehydes was tested. Correlations between the  $K_m$  of aldehydes and their hydrophobicity (expressed in terms of Hansch constant,  $\pi$ ) were found for all compounds except substituted benzaldehydes.

For FMO, significant correlations were found for OP pesticides only, albeit with a slope of 0.3, similar to the shallow slope of FMOgen. Five separate regressions were developed for ECOSAR classes in CYP. Good correlations were found for the specific chemical classes, but not for the group of diverse

chemicals ('remaining chemicals'). In the same way, the regressions for single CYP isoenzymes gave good correlations when the datasets contained mainly specific chemical classes, i.e. Anilines and Amides/Imides for CYP1A1 and Benzyl Alcohols and Esters for CYP2B4. This would suggest that lipophilicity-binding regressions for CYP isoenzymes depend on a chemical class-specific approach. Previous studies have investigated the relationship between lipophilicity and binding to CYP using homogeneous datasets. In Hansch's review on CYP [26], QSARs were developed for single experiments (single isoenzymes) on specific classes of compounds. The overall picture emerging from these models was that hydrophobic drugs are attractive targets for CYP enzymes in mammals. In Appendix C (Table C9) we reported the regressions made with the data sets in Hansch's review, which were adapted using Log  $K_{ow}$  (experimental value, if available) as sole descriptor and  $K_m$  expressed in  $\mu M$ . Among the 14 data sets, 7 gave acceptable regressions ( $n > 6$ ,  $p < 0.05$ , underlined in Table C7). In the work of Lewis and Dickins [71], QSARs were developed using  $K_m$  data collected from different enzyme assays on drugs. For a given P450 isoenzyme and for a set of substrates, a linear relationship between binding and compound lipophilicity was observed. It was described as linear free energy relationship, which is frequently encountered in biological systems. This linear relationship was not true for all compounds, possibly because of additional binding interactions involved that are not in common with those of the other substrates. Therefore, other descriptors are needed when a fairly large number of structurally diverse substrates are examined for a given P450 isoenzyme [30].

### 3.4.3 Mechanistic explanation

Lipophilicity was relevant to binding affinity for most of the substrate classes of ADH, ALDH and CYP, with the 95% CIs of the slopes (Tables C2, C3 and C5 in Appendix C) covering the value of 0.63, which is the typical slope correlating protein-water distribution (Log  $K_{pw}$ ) and Log  $K_{ow}$  [65]. The value of 0.63 is in accordance with the slopes observed in other Log  $K_{ow}$ -Log  $K_{pw}$  relationships, e.g. 0.57 (for chemicals with Log  $K_{ow}$  ranging from 2.0 to 5.1) [72] and about 0.7 [73]. A gentle slope was found for all regressions developed for FMO ( $b = 0.21-0.32$ ). If strong interactions, such as covalent or ion bonds, are important, distribution of chemicals is expected to be weakly related to their  $K_{ow}$  [65]. While the slope of the lipophilicity relationship provides an indication of the lipophilic character of the substrate binding, comparison of the intercepts indicates that at Log  $K_{ow} = 0$ ,  $1/K_m$  is about 100 times higher for ALDH than for the other enzymes family, with  $b$  of -1 and about -3, respectively.

The strength of the interactions depends on the reactions that the enzymes catalyse. ADH accepts a wide variety of substrates including exogenous

primary and secondary alcohols and oxidises them to aldehydes and ketones, respectively. ALDH metabolises endogenous and exogenous aldehydes to carboxylic alcohols (hydroxylation) [13]. FMO catalyses oxygenation of soft nucleophiles, i.e. compounds with functional groups bearing a polarisable, electron-rich centre, usually a heteroatom (such as nitrogen, sulphur and phosphorus) in organic compounds [74]. The poor correlation found for FMO could be attributed to its catalytic cycle, which is different with respect to the other enzymes [58]. FMO is a flavin protein containing a single FAD, which is first reduced and then reacts with molecular oxygen to form a peroxy-flavin (FADOOH), which can subsequently react with the substrate. The nucleophilic attack on the FADOOH results in the transfer of 1 atom of molecular oxygen on the substrate. The access to the FADOOH intermediate could be better predicted by descriptors such as electronic properties rather than lipophilicity. CYP is involved in the metabolism (primarily oxidative) of a vast number and wide structural variety of compounds [49]. In an extensive study on CYP3A4 [75], among the various types of mediated reactions, the best lipophilicity- $K_m$  correlation was achieved for carbon hydroxylation, while no or little correlations were seen for N-, S-oxidation and other reactions. Also in our study, hydroxylation (mediated by ALDH) gave the best regressions, while for N-, S-oxidation (mediated by FMO) a poor correlation was found between  $K_m$  and  $K_{ow}$ .

#### 3.4.4 Application

The regressions obtained in the present study relate the enzyme binding with  $\text{Log } K_{ow}$ , the descriptor which is commonly used in bioaccumulation models. Information on both  $K_m$  and  $V_{max}$  is essential for the extrapolation from *in vitro* to *in vivo* metabolism, required for risk assessment. In fact, for reactions that exhibit Michaelis-Menten kinetics and on condition of non-saturating substrate concentration, the ratio between  $V_{max}$  and  $K_m$  provides an estimation of the intrinsic clearance ( $CL_{int}$ ) [19, 76]. This parameter, which is a measure of enzyme activity towards a compound, can be extrapolated to equivalent whole-body metabolic rate [77]. Yet, in order to apply these regressions to predict whole-body metabolic rates, improvements are needed at various points. Firstly, the explained variance ( $r^2$ ) of the present regressions can be increased by extending the number of descriptors included, such as hydrogen-bond descriptors. In addition, other investigations are required to predict  $V_{max}$ , in order to understand also the processes that control the catalytic step of metabolism.

## **Appendices**

Appendix B provides original  $K_m$  data.

Appendix C provides regressions including 95% CI intervals, Log  $K_{ow}$  and Log  $(1/K_m)$  ranges, regressions for rat data (ALDH), regressions using Log  $D_{7.4}$  values and regressions for single CYP experiments, as well as additional tables listing substituted benzaldehydes and DTC, OP and CM pesticides, with their general chemical structure.

## **Acknowledgments**

Financial support by the European Union through the Environmental ChemOinformatics (ECO) Project (FP7-PEOPLE-ITN-2008, no. 238701) is gratefully acknowledged.





## Chapter 4

# **Mechanistically-based QSARs to describe metabolic constants in mammals**

Alessandra Pirovano

Mark A.J. Huijbregts

Ad M.J. Ragas

Karin Veltman

A. Jan Hendriks

*Alternatives to Laboratory Animals* (2014), 42(1), 59–69

## 4.1 Introduction

The bioaccumulation potential of chemicals in organisms is a vital element in environmental risk assessment [1]. The accumulation of a chemical is the result of a series of physiological and physical processes: absorption, distribution, metabolism and excretion (ADME). Metabolism, also referred to as biotransformation in the case of xenobiotics [1], occurs via enzymatic reactions involving two processes. Firstly, the chemical needs to reach the enzyme and bind to it; secondly, a catalytic reaction has to take place. The latter process is described by the maximum rate of reaction ( $V_{\max}$ ) at saturating substrate concentration [25]. Alternatively,  $V_{\max}$  can be expressed as turnover number ( $k_{\text{cat}}$ , with units of  $\text{time}^{-1}$ ), which represents the number of substrate molecules converted into product per enzyme molecule per time, when the enzyme is saturated with substrate [78]. The other parameter used to characterise an enzymatic reaction is the Michaelis-Menten constant  $K_m$ , which is the substrate concentration at half  $V_{\max}$ .  $K_m$  is equal to the ratio  $(k_{\text{cat}} + k_{-1})/k_1$ , where  $k_{-1}$  and  $k_1$  are constants, respectively, for breakdown and formation of the complex enzyme-substrate (ES) [24]. If  $k_{\text{cat}}$  is smaller than  $k_{-1}$ ,  $K_m$  is assumed to be equal to the dissociation constant  $K_d$  for the ES complex. In this case,  $1/K_m$  reflects the affinity of the enzyme for its substrate: a low  $K_m$  (or high  $1/K_m$ ) corresponds to high binding affinity. For reactions that exhibit Michaelis-Menten kinetics and at non-saturating substrate concentrations, the ratio  $V_{\max}/K_m$  provides an estimation of the intrinsic clearance ( $CL_{\text{int}}$ ) [19, 76].  $CL_{\text{int}}$ , which is a measure of enzyme activity toward a compound, can be extrapolated to an equivalent whole-body metabolic rate, required for risk assessment [77].

Several studies [26, 27] have shown the importance of Quantitative Structure-Activity Relationships (QSARs) for the investigation of  $K_m$  and  $V_{\max}$ , most of which focused on drugs oxidised by cytochrome P450 (CYP). The binding to the enzyme, represented by  $1/K_m$ , was shown to be mainly related to compound hydrophobicity [25, 28], probably due to desolvation effects, although electronic and geometric factors, such as polarity and size, can also be important [27]. The rate appears to be influenced by electronic properties, such as frontier orbital energies or hydrogen bonding properties [29-31]. In fact, catalytic processes are characterised by cleavage and formation of covalent bonds [25]. However, the above-mentioned studies focussed on particular series of P450 substrates, implying applicability only for specific combinations of chemicals and P450 enzymes. Recently, Pirovano et al. [79] studied the relationships between  $1/K_m$  and hydrophobicity, i.e. the octanol-water partitioning coefficient ( $K_{\text{ow}}$ ), for a broader set of chemicals and oxidising enzymes in mammals. The chemicals investigated were xenobiotics such as alcohols, aldehydes, drugs and pesticides. The enzymes examined, in addition

to CYP, were alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH) and flavin-containing monooxygenase (FMO).

In the present study, we extended our analysis to other descriptors, which were chosen on the basis of mechanistic considerations. Furthermore, we did not only investigate descriptors for  $1/K_m$ , but also for  $V_{max}$ . The aim of the current study was to develop QSARs with  $\text{Log}(1/K_m)$  and  $\text{Log } V_{max}$  as endpoints for ADH, ALDH, FMO and CYP enzymes in mammals. General linear models were built with descriptors related to partitioning, as well as geometric and electronic properties of the substrates.

## 4.2 Materials and methods<sup>1</sup>

### 4.2.1 Experimental dataset

#### *Data collection*

$K_m$  and catalytic reaction rates (expressed either as  $V_{max}$  or  $k_{cat}$ ) were taken from peer-reviewed articles. We considered the following enzymes: alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), flavin-containing monooxygenase (FMO) and cytochrome P450 (CYP). For ADH and ALDH, data were taken from the BRENDA enzyme database [57] (BRaunschweig ENzyme DATabase, <http://www.brenda-enzymes.org>). Metabolic constants for FMO were taken from a review [58] and references contained therein. Data for CYP were sourced from other reviews [26, 59, 60]. All data extracted from the BRENDA database and the reviews were checked in the original papers. Constants measured for mammals in *in vitro* assays of purified, non-recombinant, hepatic enzymes were selected. For each value, we recorded the species and the enzyme for which it was measured and the experimental conditions such as pH and temperature. Rate values were not reported in one article on ALDH [80] and six articles on FMO [81-86], in which only  $K_m$  values were measured for a total of 5 and 75 compounds, respectively. The substrates collected are mainly drugs and compounds found in the environment.

SMILES (Simplified Molecular Input Line Entry System) strings [61] and CAS (Chemical Abstract Service) numbers were obtained from the ChemSpider website (<http://www.chemspider.com/>). Each compound was assigned to a relevant chemical class using ECOSAR v 1.0, a program present in the EPI Suite of the US Environmental Protection Agency (EPA) (<http://www.epa.gov>) [87].

---

<sup>1</sup> In the original paper, a shortened version of the Materials and Methods section was present. The extended version present in this thesis was reported in the Supporting Information.

The data collected can be found in Appendix B (Table B1), with the references to the original papers.

#### *Data treatment*

Michaelis constants ( $K_m$ ) were expressed in  $\mu\text{M}$ . Since catalytic rates were reported in heterogeneous units and with different constants (i.e. as  $V_{\max}$  or as  $k_{\text{cat}}$ ), it was necessary to standardise the data. We expressed all rates as  $V_{\max}$ , using  $\mu\text{mol min}^{-1} \text{mg}_{\text{PROT}}^{-1}$  as units. For CYP enzymes, assays were performed isolating microsomal fractions and inducing the activity of the P450 isoenzyme of interest by treating the animals with various agents, such as Phenobarbital for CYP2B1 in rat [88];  $V_{\max}$  was then referred to the microsomal protein weight, i.e.  $\text{mg}_{\text{PROT}} = \text{mg}_{\text{MICR PROT}}$ . For the other enzymes,  $V_{\max}$  was referred to the weight of the enzyme being studied, i.e.  $\text{mg}_{\text{PROT}} = \text{mg}_{\text{ENZ}}$ , as the assays were performed with isolated and purified liver enzymes. The rates expressed as  $k_{\text{cat}}$  were transformed into  $V_{\max}$  values. For ADH, ALDH and FMO, we derived  $V_{\max}$  (expressed as  $\mu\text{mol min}^{-1} \text{mg}_{\text{ENZ}}^{-1}$ ) dividing  $k_{\text{cat}}$  ( $\text{min}^{-1}$ ) by the molecular weight of the enzyme ( $M_r$ ,  $\text{mg}_{\text{ENZ}} \mu\text{mol}^{-1}$ ). For CYP, we transformed  $k_{\text{cat}}$  ( $\text{min}^{-1}$ ) into  $V_{\max}$  values (expressed as  $\mu\text{mol min}^{-1} \text{mg}_{\text{MICR PROT}}^{-1}$ ) multiplying the former by the specific content of the enzyme ( $E$ ,  $\mu\text{mol mg}_{\text{MICR PROT}}^{-1}$ ) [29]. If  $M_r$  or  $E$  values were not reported in the paper where we collected  $k_{\text{cat}}$ , we used average values coming from other studies. The operations performed on the data are reported in detail in Appendix B, Table B2. The  $V_{\max}$  values expressed in  $\mu\text{mol min}^{-1} \text{mg}_{\text{PROT}}^{-1}$  and used in this study are reported in Appendix B, Table B1, together with the original rate values.

Michaelis constants ( $K_m$ ) and maximum rates ( $V_{\max}$ ) of different substrates were combined into 4 datasets, one for each enzyme family. 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 of the experimental  $1/K_m$  or  $V_{\max}$  values, as well as the geometric standard deviation.

#### *4.2.2 Descriptors and QSAR models*

##### *Descriptor calculation and selection*

We compiled a list of physicochemical descriptors based on mechanistic considerations. We anticipated  $1/K_m$  and  $V_{\max}$  to be related to the partitioning, geometric and electronic properties of the substrates of P450 [29, 59, 66, 89]. Therefore, we collected the descriptors (18 in total) used in the QSARs for Log ( $1/K_m$ ) or Log  $V_{\max}$  in the above-mentioned studies. We hypothesised that they could be applied to all four enzyme classes, as they were among the descriptors commonly used to describe biological responses to xenobiotics [90]. The descriptors were computed with Chemaxon (<http://www.chemaxon.com>) through the OCHEM platform [91]

(<http://ochem.eu>) and with the semi-empirical molecular orbital program MOPAC2009 [92] (Hamiltonian AM1) using the software Vega ZZ [93] v2.4.0 (<http://vegazz.net>). For the calculation of all descriptors, the molecular conformations were optimized with MOPAC. A correlation matrix was calculated on all compounds as a first screening to detect collinear descriptors, i.e. descriptors with correlation coefficients (R) higher than 0.8 or lower than -0.8. Among the collinear descriptors, we retained the one that we considered easier to interpret mechanistically.

The final set of descriptors is reported in Table 4.1, together with the software used to compute them. The partitioning was expressed with the octanol water partitioning coefficient of the uncharged molecule (logP). The geometrical descriptors of the chemicals were molecular area (A), i.e. length times width, ratio of molecular length to molecular width (l/w) and ratio of the area of the molecule to the square of depth ( $a/d^2$ ). Length, width and depth of a molecule represent molecular dimensions measured orthogonally relative to the main molecular plane [94]. The electronic parameters were the strongest acidic and strongest basic  $pK_a$  ( $apK_a1$  and  $bpK_a1$ , respectively), hydrogen bond donor and acceptor (HBD and HBA, respectively), dipole moment ( $\nu$ ), final heat of formation ( $H_f$ ), energy of the highest occupied molecular orbital and energy of the lowest unoccupied molecular orbital ( $E_{HOMO}$  and  $E_{LUMO}$ , respectively) and the difference between the frontier orbital energy levels ( $\Delta E_{L-H} = E_{LUMO} - E_{HOMO}$ ). The descriptors were auto-scaled to zero-mean and unit-variance to ensure equal contribution of all variables in the models.

### *Model development*

General linear models (GLM) were developed for  $\text{Log}(1/K_m)$  and  $\text{Log } V_{\max}$  with the software R v.2.15.1 [95] (<http://www.R-project.org>). We used the R package 'bestglm' [96] to select the best subset of variables for the linear regression after an exhaustive search, i.e. all possible combinations of descriptors were generated and tested by the algorithm. In order to avoid overfitting, we set the maximum number of variables to be included in the subsets at 6. It is generally recommended that the ratio of number of compounds to the number of descriptors in the QSAR should be at least 5:1 [97]. The best model was then chosen based on the Akaike's Information Criterion (AIC). The AIC is a trade-off between a good fit to the model (measured by the likelihood) and a penalty for complexity (calculated using the number of parameters). The model with the lowest AIC is interpreted as the best model. We performed a final check for collinearity of the descriptors in the individual QSARs using variance inflation factors (VIFs). We used the R package 'car' [98] to calculate the VIFs for the variables included in each QSAR in order to check if they were collinear. The threshold for collinearity was

VIF>3 [99]. If all variables had VIFs<3, the QSAR was accepted; otherwise, the variable with the highest VIF was removed from the dataset and the 'bestglm' method was performed again. The VIF values were then recalculated and this procedure was repeated until all VIF values were smaller than the threshold [100].

The models were cross-validated with the leave-one-out (LOO) procedure using WEKA v.3.6.7 [101] (<http://www.cs.waikato.ac.nz>). With the LOO cross validation, a single observation is removed from the original dataset, and the remaining observations are used as training data, in such a way that each observation is removed only once. Then one model is developed for each data set, and the response values of the removed observations are predicted from these models.

For each model, we report the coefficient of determination ( $R^2$ ) and the Root Mean Squared Error (RMSE) as measures of the fitting. The adjusted coefficient of determination ( $R^2_{adj}$ ) is shown in order to adjust the  $R^2$  value for the number of explanatory variables in the model. The fitting of the models is also evaluated based on the p-value from the F-test (p). We report the LOO cross-validated  $R^2$  ( $Q^2_{LOO}$ ) and RMSE ( $RMSE_{LOO}$ ) to assess the predictive power of the models. The formulas of these coefficients are presented in Appendix B. In the equations of the QSARs, we show the standardised coefficients of the variables (i.e. the regression coefficients that do not depend on the units and that were obtained using the auto-scaled descriptors) together with their errors and p-values.

#### *Additional regressions*

In our previous work on the relationship between  $1/K_m$  and lipophilicity [79], we observed two groups of compounds that were outliers: 22 substituted benzaldehydes for ALDH (listed in the Appendix C, Table C6) and 52 'non specific' chemicals for CYP (mainly Neutral Organics, according to the ECOSAR classification). Therefore, in this work we also investigated the possible influence of these classes of compounds in the QSARs. We developed two additional sets of QSARs for both ALDH and CYP: one with all compounds except the group of outliers ( $ALDH_1$  and  $CYP_1$ ) and one with only the group of outliers ( $ALDH_2$  and  $CYP_2$ ). For the QSARs with the 22 substituted benzaldehydes, the maximum number of variables to be selected by the algorithm was set to 4, due to the relatively low number of compounds. We also developed an overall regression for  $\text{Log}(1/K_m)$ , merging all data from the 4 datasets and adding a qualitative variable called "Enzyme" with four categories (ADH, ALDH, FMO, CYP) representing the enzyme group of the data point.

Table 4.1. Descriptors used to develop the QSARs.

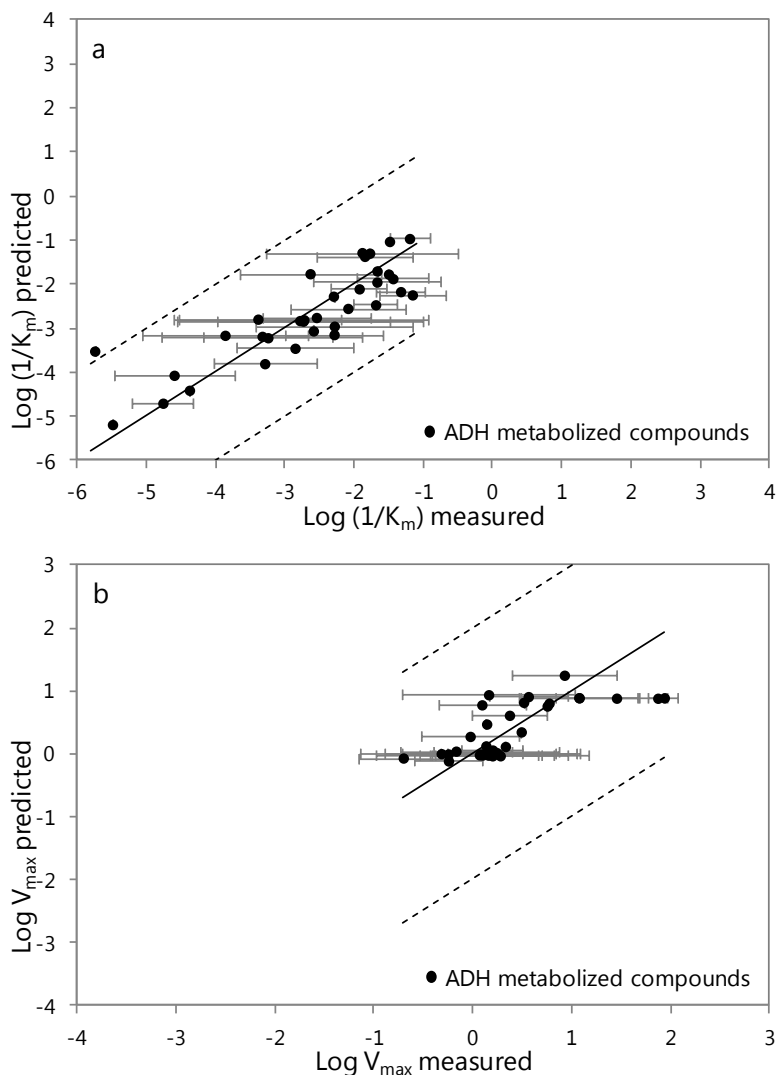
Symbol	Units	Description	Type	Software
logP	[-]	Calculated octanol water partitioning coefficient	Partitioning	Chemaxon
A	[Å <sup>2</sup> ]	Van der Waals surface area, calculated at pH 7.4	Geometric	Chemaxon
a/d <sup>2</sup>	[-]	area/depth <sup>2a</sup>	Geometric	MOPAC
l/w	[-]	length/width <sup>a</sup>	Geometric	MOPAC
apK <sub>a</sub> 1	[-]	Strongest acidic pK <sub>a</sub>	Electronic	Chemaxon
bpK <sub>a</sub> 1	[-]	Strongest basic pK <sub>a</sub>	Electronic	Chemaxon
HBD	[-]	Hydrogen bond donor, calculated at pH 7.4	Electronic	Chemaxon
HBA	[-]	Hydrogen bond acceptor, calculated at pH 7.4	Electronic	Chemaxon
ν	[Debye]	Dipole moment	Electronic	MOPAC
E <sub>HOMO</sub>	[eV]	Energy of the highest occupied molecular orbital (HOMO)	Electronic	MOPAC
E <sub>LUMO</sub>	[eV]	Energy of the lowest unoccupied molecular orbital (LUMO)	Electronic	MOPAC
ΔE <sub>L-H</sub>	[eV]	ΔE <sub>L-H</sub> = E <sub>LUMO</sub> - E <sub>HOMO</sub>	Electronic	MOPAC
H <sub>f</sub>	[kcal/mol]	Final heat of formation	Electronic	MOPAC

<sup>a</sup> Length, width and depth of a molecule represent molecular dimensions measured orthogonally relative to the main molecular plane (35).

### 4.3 Results

The QSARs developed for Log (1/K<sub>m</sub>) and Log V<sub>max</sub> are presented in Tables 4.2 and 4.3, respectively, with the standardised regression coefficients (i.e. the regression coefficients that do not depend on the units and were obtained by using the auto-scaled descriptors). The non-standardised regression coefficients and the overall regression for Log (1/K<sub>m</sub>) and Log V<sub>max</sub> can be found in the Appendix D, Tables D1-D2. As an example, Figure 4.1 represents the measured versus the predicted values for Log (1/K<sub>m</sub>) and Log V<sub>max</sub> for ADH.

Figure 4.1 Measured versus predicted values for a)  $\text{Log } (1/K_m)$  and b)  $\text{Log } V_{\max}$ , for compounds metabolised by ADH in mammals. The solid lines indicate the 1:1 bisector and the dashed lines indicate  $\pm 2$  Log units error. Laboratory measurements (dots) for each compound: Log transformed geometrical mean of a)  $1/K_m$  [ $\mu\text{M}^{-1}$ ] and b)  $V_{\max}$  [ $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}_{\text{PROT}}^{-1}$ ], with the geometric standard deviation (horizontal bar).





#### 4.3.1 Log (1/ $K_m$ )

Significant correlations ( $p < 0.05$ ) were obtained for all QSARs for Log 1/ $K_m$  (Table 4.2), whose  $R^2_{adj}$  and  $Q^2_{LOO}$  varied from 0.37 to 0.74 and from 0.30 to 0.72, respectively. The most common descriptors were area (A), octanol-water partitioning coefficient (logP) and difference between frontier orbital energies ( $\Delta E_{L-H}$ ). The area had positive regression coefficients, ranging from 0.25 to 1.02. The coefficients of logP and  $\Delta E_{L-H}$  had positive and negative signs, respectively, in all cases, except for ALDH<sub>2</sub> (QSAR with only the 22 substituted benzaldehydes) and ALDH (only for  $\Delta E_{L-H}$ ). These 3 descriptors were the most important ones, i.e. with the highest standardised coefficients, in most of the QSARs: the area for ALDH, ALDH<sub>2</sub> and CYP<sub>1</sub> (QSAR without the 'remaining chemicals'); logP for ADH, ALDH<sub>1</sub> (QSAR without the 22 substituted benzaldehydes);  $\Delta E_{L-H}$  for CYP. The hydrogen bond acceptor (HBA) had the highest standardised regression coefficient (-0.42) in the QSAR for FMO.

#### 4.3.2 Log $V_{max}$

Correlations significant at the 0.05 level were obtained for all QSARs for Log  $V_{max}$  (Table 4.3). The goodness of fit and the internal predictivity were lower for Log  $V_{max}$ , if compared to Log (1/ $K_m$ ), with  $R^2_{adj}$  and  $Q^2_{LOO}$  varying from 0.17 to 0.48 and from 0.12 to 0.41, respectively. The most common descriptor, appearing in six out of eight QSARs, was the dipole moment ( $\nu$ ), with coefficients ranging from -0.42 to 0.36. It was also the most important descriptor in the QSARs for ALDH<sub>1</sub> and CYP<sub>1</sub>. The area (A) featured in four QSARs with a positive regression coefficient; it had the highest standardised coefficient (0.37) in the QSARs for ALDH and ALDH<sub>2</sub>. LogP occurred in three QSARs with a negative regression coefficient and, with a standardised coefficient of -0.27, it was the most important descriptor for ALDH<sub>1</sub>, together with the dipole moment. Among the other descriptors,  $apK_{a1}$ , HBA and  $E_{LUMO}$  had the highest correlation coefficients for FMO (-0.15), CYP<sub>2</sub> (-0.29) and ADH (-0.44), respectively.  $H_f$  was the most important descriptor for CYP, with a standardised coefficient of 0.21.

Table 4.2a. Log (1/K<sub>m</sub>): Variables selected and their standardised regression coefficients (for symbols see Table 4.1). The K<sub>m</sub> values were expressed as μM. The most important descriptor of each regression is shown in bold.

Enzyme	logP	A	a/d <sup>2</sup>	I/w	apK <sub>a</sub> 1	bpK <sub>a</sub> 1	HBD	HBA	v	E <sub>HOMO</sub>	E <sub>LUMO</sub>	ΔE <sub>L-H</sub>	H <sub>f</sub>	Interc.
ADH	<b>0.98</b> (±0.16)		-0.24 (±0.13) <sup>a</sup>		-0.34 (±0.15)		-0.25 (±0.15) <sup>a</sup>				-0.36 (±0.13)			-2.66 (±0.12)
ALDH	0.64 (±0.18)	<b>0.82</b> (±0.19)					-0.30 (±0.14)	0.54 (±0.20)				0.51 (±0.21)	0.39 (±0.15)	-0.18 (±0.13) <sup>a</sup>
FMO		0.25 (±0.08)						<b>-0.42</b> (±0.07)				-0.30 (±0.07)		-1.99 (±0.06)
CYP		0.30 (±0.07)		0.20 (±0.07)	-0.18 (±0.07)				-0.10 (±0.07) <sup>a</sup>			<b>-0.36</b> (±0.07)		-2.73 (±0.06)
Additional regressions														
ALDH <sub>1</sub>	<b>0.65</b> (±0.15)	0.63 (±0.14)		0.24 (±0.12) <sup>a</sup>	-0.36 (±0.11)	0.27 (±0.12)	-0.28 (±0.12)							-0.12 (±0.11) <sup>a</sup>
ALDH <sub>2</sub>	-0.57 (±0.34) <sup>a</sup>	<b>1.02</b> (±0.31)										0.53 (±0.31) <sup>a</sup>		-0.34 (±0.27) <sup>a</sup>
CYP <sub>1</sub>	0.32 (±0.07)	<b>0.36</b> (±0.07)			-0.09 (±0.06) <sup>a</sup>		-0.09 (±0.06) <sup>a</sup>				-0.25 (±0.07)		0.12 (±0.06)	-2.80 (±0.05)
CYP <sub>2</sub>				<b>0.53</b> (±0.10)	-0.13 (±0.10) <sup>a</sup>							-0.32 (±0.10)		-2.65 (±0.09)

<sup>a</sup> The probability (p) value of the coefficient is greater than 0.05.

Table 4.2b. Regression statistics for the QSARs in Table 4.2a.

Enzyme	n	R <sup>2</sup>	R <sup>2</sup> <sub>adj</sub>	RMSE	p	Q <sup>2</sup> <sub>loo</sub>	RMSE <sub>loo</sub>
ADH	34	0.73	0.68	0.62	<0.01	0.57	0.80
ALDH	77	0.56	0.52	1.06	<0.01	0.47	1.17
FMO	149	0.39	0.37	0.63	<0.01	0.35	0.79
CYP	121	0.40	0.37	0.68	<0.01	0.30	0.73
<b>Additional regressions</b>							
ALDH <sub>1</sub>	55	0.77	0.74	0.74	<0.01	0.72	0.82
ALDH <sub>2</sub>	22	0.53	0.46	1.16	<0.01	0.36	1.40
CYP <sub>1</sub>	65	0.73	0.71	0.39	<0.01	0.60	0.48
CYP <sub>2</sub>	56	0.52	0.50	0.68	<0.01	0.45	0.74

Table 4.3. Log  $V_{\max}$ : Variables selected and their standardised regression coefficients (for symbols see Table 4.1), together with the regression statistics. The  $V_{\max}$  values were expressed as  $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}_{\text{PROT}}^{-1}$ . The most important descriptor of each regression is shown in bold.

Enzyme	logP	A	$a/d^2$	I/w	apK <sub>a</sub> 1	bpK <sub>a</sub> 1	HBD	HBA	$\nu$	E <sub>HOMO</sub>	E <sub>LUMO</sub>	$\Delta E_{\text{L-H}}$	H <sub>f</sub>	Interc.
ADH								-0.25 ( $\pm 0.09$ )	0.36 ( $\pm 0.09$ )		<b>-0.44</b> ( $\pm 0.08$ )			0.38 ( $\pm 0.07$ )
ALDH	-0.35 ( $\pm 0.09$ )	<b>0.37</b> ( $\pm 0.11$ )			0.10 ( $\pm 0.07$ ) <sup>a</sup>				-0.17 ( $\pm 0.09$ ) <sup>a</sup>		0.23 ( $\pm 0.09$ )		0.16 ( $\pm 0.09$ ) <sup>a</sup>	-0.44 ( $\pm 0.07$ )
FMO				-0.07 ( $\pm 0.03$ )	<b>-0.15</b> ( $\pm 0.03$ )	0.04 ( $\pm 0.03$ ) <sup>a</sup>		0.13 ( $\pm 0.03$ )	0.06 ( $\pm 0.03$ ) <sup>a</sup>	0.10 ( $\pm 0.03$ )				-0.19 ( $\pm 0.03$ )
CYP			0.14 ( $\pm 0.06$ )		0.18 ( $\pm 0.06$ )		-0.09 ( $\pm 0.06$ ) <sup>a</sup>		-0.16 ( $\pm 0.05$ )	0.19 ( $\pm 0.06$ )			<b>0.21</b> ( $\pm 0.06$ )	-1.45 ( $\pm 0.05$ )
ALDH <sub>1</sub>	<b>-0.27</b> ( $\pm 0.09$ )	0.25 ( $\pm 0.10$ )		0.15 ( $\pm 0.08$ ) <sup>a</sup>			-0.17 ( $\pm 0.07$ )		<b>-0.27</b> ( $\pm 0.09$ )					-0.29 ( $\pm 0.06$ )
ALDH <sub>2</sub>		<b>0.37</b> ( $\pm 0.16$ )					0.34 ( $\pm 0.16$ )							-0.81 ( $\pm 0.15$ )
CYP <sub>1</sub>	-0.19 ( $\pm 0.07$ )		0.34 ( $\pm 0.09$ )		0.19 ( $\pm 0.09$ )			0.23 ( $\pm 0.10$ )	<b>-0.42</b> ( $\pm 0.11$ )		0.24 ( $\pm 0.10$ )			-1.63 ( $\pm 0.07$ )
CYP <sub>2</sub>		0.26 ( $\pm 0.10$ )		0.18 ( $\pm 0.06$ )	-0.11 ( $\pm 0.06$ ) <sup>a</sup>		0.14 ( $\pm 0.07$ ) <sup>a</sup>	<b>-0.29</b> ( $\pm 0.08$ )				0.19 ( $\pm 0.09$ )		-1.23 ( $\pm 0.06$ )

<sup>a</sup> The probability (p) value of the coefficient is greater than 0.05.

Table 4.3b. Regression statistics for the QSARs in Table 4.3a.

Enzyme	n	R <sup>2</sup>	R <sup>2</sup> <sub>adj</sub>	RMSE	p	Q <sup>2</sup> <sub>loo</sub>	RMSE <sub>loo</sub>
ADH	33	0.53	0.48	0.40	<0.01	0.41	0.45
ALDH	74	0.25	0.19	0.55	<0.01	0.15	0.60
FMO	94	0.30	0.25	0.26	<0.01	0.20	0.28
CYP	121	0.26	0.22	0.55	<0.01	0.17	0.59
<b>Additional regressions</b>							
ALDH <sub>1</sub>	52	0.25	0.17	0.44	0.02	0.12	0.51
ALDH <sub>2</sub>	22	0.29	0.22	0.64	0.04	0.12	0.74
CYP <sub>1</sub>	65	0.37	0.30	0.54	<0.01	0.24	0.60
CYP <sub>2</sub>	56	0.43	0.36	0.40	<0.01	0.24	0.47

## 4.4 Discussion

### 4.4.1 Regressions

In this study, QSAR models were developed for  $\text{Log}(1/K_m)$  and  $\text{Log } V_{\max}$  of four groups of mammalian enzymes. We used relevant physicochemical descriptors reflecting hydrophobic, geometric and electronic properties of the chemicals. Common features were found within the QSARs for  $\text{Log}(1/K_m)$  and  $\text{Log } V_{\max}$ , despite the different reaction types of the four enzymes considered.  $\text{Log}(1/K_m)$  was largely controlled by hydrophobicity ( $\log P$ ), as well as area ( $A$ ) and frontier orbital energy ( $\Delta E_{L-H}$ ), while the rate ( $V_{\max}$ ) was mainly influenced by electronic parameters, such as dipole moment ( $\nu$ ), hydrogen bonding properties (HBD and HBA) and energy of the lowest occupied molecular orbital ( $E_{\text{LUMO}}$ ). The difference in the molecular properties controlling  $\text{Log}(1/K_m)$  and  $\text{Log } V_{\max}$  was expected from the nature of the processes underlying these two constants. The inverse of  $K_m$  is usually assumed to be equal to the affinity constant for enzyme binding, which is generally a desolvation process; thus, it is controlled mainly by hydrophobicity. Yet, this equivalence is valid only if the enzymatic process is composed of two steps - formation of the ES complex and successive catalysis - and if the latter is lower than the dissociation of the substrate from the enzyme. The  $V_{\max}$  represents the catalytic process, which is characterised by the cleavage and formation of covalent bonds; thus, it is more influenced by electronic properties of the substrates [25].

The variability explained by the QSARs ranged from 20% to 70% ( $R^2_{\text{adj}}$  in Tables 4.2 and 4.3). The correlations improved substantially for  $\text{Log}(1/K_m)$  by leaving out distinct substance groups such as substituted benzaldehydes. Weak correlations may indicate that the underlying catalytic reactions are complex and only partly related to the physicochemical descriptors chosen [102]. The fit of the QSARs could be improved by using theoretical molecular descriptors, i.e. calculated by mathematical formulae or computational algorithms [103], which are able to represent other aspects of molecular structures, such as topological indices and functional group counts. Yet, we did not include these descriptors in the present paper, because the objective was to allow for the mechanistic interpretation of the QSARs.

The QSARs in the present work had lower  $R^2$  values in comparison to the QSARs for CYP developed in other studies, whose  $R^2$  values were around 0.8-0.9 [26, 27, 29, 59]. Yet, the latter datasets typically included homologous series of about ten structurally-related compounds, metabolised by a given isoenzyme in one mammalian species. Thus, those models are applicable only to very specific combinations of compounds, isoenzymes and species, for which a similar behaviour can be expected.

The datasets consisted of compounds assigned to ECOSAR classes and known to be substrates of the enzymes considered in this study. The applicability domains of the QSARs are defined by the range (min and max) of the values of the descriptors used to build the model [104], which are reported in Tables D3 and D4 in Appendix D for  $\text{Log}(1/K_m)$  and  $\text{Log } V_{\max}$ , respectively.

The experimental data come from different laboratories, often employing different protocols [67], e.g. for pH and temperature conditions, which can affect enzyme activity [78]. In addition, the rates were reported in the papers either as  $V_{\max}$  or  $k_{\text{cat}}$  values. The latter were transformed into  $V_{\max}$  (Appendix B, Table B2) by using the conversion factors reported in the papers from which we collected  $k_{\text{cat}}$ , when available; otherwise, we used average values obtained in other studies. Consequently, the input data are subject to variation, implying uncertainty in the QSARs. Furthermore, we merged data measured for different mammalian species (human, horse, rat, mouse, pig and rabbit) and isoenzymes (i.e. any of the several forms of an enzyme, all of which catalyse the same reaction but are characterised by different properties). This can be another source of unexplained variation; however, the focus of the present work was on general features in the metabolic process.

We built four general QSARs each for  $\text{Log}(1/K_m)$  and  $\text{Log } V_{\max}$ , one for every enzyme. In our previous study [79] on the relationships between hydrophobicity and  $\text{Log}(1/K_m)$ , we found an improvement of the regressions after the removal of two groups of influential chemicals: 22 substituted benzaldehydes for ADH and 56 ‘remaining chemicals’ (chemicals belonging to non-specific ECOSAR classes, mainly Neutral Organics) for CYP. Hence, in the present study, we developed four additional QSAR sub-models each for  $\text{Log}(1/K_m)$  and  $\text{Log } V_{\max}$ , one without and one with only the groups of influential chemicals. For ALDH, the fitting increased with respect to the general QSAR only for the sub-model built for  $\text{Log}(1/K_m)$  excluding the substituted benzaldehydes ( $\text{ALDH}_1$ ). For both endpoints, the most important descriptor was the area for the substituted benzaldehydes and  $\log P$  for the other aldehydes. For CYP, this subdivision lead to QSAR sub-models with improved fitting for both  $\text{Log}(1/K_m)$  and  $\text{Log } V_{\max}$ , although for the latter the  $Q^2_{\text{LOO}}$  values were low (around 0.2). It appears that the enzymatic constants can be dependent on chemical classes. The ‘remaining chemicals’ for CYP may have different abilities to fit onto and interact with the enzyme active site.

#### 4.4.2 Mechanistic explanation

The QSARs developed in this work were generally in line with previous studies on enzyme metabolism, mainly concerning P450 enzymes [27]. In the following paragraphs, the influencing descriptors in the QSARs are explained in relation to the catalytic cycles of the enzymes. Liver ADH catalyses the reversible

transformation of alcohols to the corresponding aldehydes or ketones. ALDH enzymes oxidise a wide range of aldehydes to their corresponding carboxylic acids [105]. FMO oxygenates various xenobiotics, such as pesticides and drugs, containing a nucleophilic heteroatom (usually sulphur and nitrogen) [58]. The oxygen abstraction takes place before binding via a nucleophilic attack by the substrate. The CYP enzymes usually catalyse mono-oxygenase reactions involving the insertion of an oxygen atom into a substrate [29].

The hydrophobicity ( $\log P$ ) featured in many QSARs for  $\log (1/K_m)$ , for which it had a positive correlation coefficient, with the exception of the QSAR for 'substituted benzaldehydes'. The increase of  $1/K_m$  with compound hydrophobicity is likely to indicate the importance of weak interactions such as substrate binding via desolvation processes, i.e. displacement of water molecules due to the binding of the substrate in the active site [106]. The different behaviour of substituted benzaldehydes was observed in our previous work relating  $\log (1/K_m)$  to compound hydrophobicity (relationships shown in Appendix D, Tables D5-D6). In the work of Klyosov [70], correlations between the  $K_m$  of aldehydes and their hydrophobicity (expressed in terms of Hansch constant,  $\pi$ ) were found for all aldehydes tested except substituted benzaldehydes. In our QSARs for  $\log V_{max}$ ,  $\log P$  featured only in three QSARs, which is in accordance with the common understanding that rates are not likely to be influenced by partitioning properties. In addition,  $\log P$  had a negative coefficient for  $\log V_{max}$ , indicating that hydrophobicity disfavors the catalysis of the substrates.

Geometric properties of the substrates were included in several QSARs, the most frequent being the molecular area ( $A$ ), always with a positive regression coefficient. The area was often the most important descriptor for  $\log (1/K_m)$ , and its contribution might be explained in two possible ways. First, larger dimensions increase the possibility of interactions with the binding site, which is an effect purely related to size. In addition, the area can be an indicator of compound hydrophobicity, as large molecules are often more hydrophobic. Thus, in the QSARs for  $\log (1/K_m)$ , the presence of the area reconfirmed the hydrophobic nature of the binding sites of the enzymes. For FMO and CYP, the area featured in the QSARs for  $1/K_m$ , but the most important descriptors were related to electronic properties. In these cases,  $1/K_m$  may not be an indicator of binding, as it describes stronger interactions. The catalytic mechanism of FMO involves a nucleophilic attack, which takes place before binding [58]. CYP enzymes have a catalytic mechanism with many steps occurring between binding and substrate oxygenation [49]. It was shown that  $K_m$  values may be sensitive to kinetic perturbations at catalytic steps taking place after substrate binding; thus,  $1/K_m$  values may not be good approximations of affinity constants [107]. The electronic descriptors related to protonation ( $\text{ap}K_a1$  and



bpK<sub>a</sub>1) featured in many QSARs, especially the acidic dissociation constant, which had negative regression coefficients for Log (1/K<sub>m</sub>). This means that 1/K<sub>m</sub> is higher for more acidic compounds (i.e. with lower pK<sub>a</sub>). The ionisation constant was a relevant factor also in QSARs for microbial biodegradation [108], due to the importance of protonation for enzyme-substrate interactions, as well as for penetration of the compound through the lipid bilayer. Electronic descriptors, such as HBD, HBA and dipole moment ( $\nu$ ), featured quite often especially in the QSARs for Log V<sub>max</sub>. This indicates that hydrogen bonding and polarity may play a significant role in the substrate-enzyme interactions.

In our study, we included frontier orbital parameters associated with metabolic properties: the energy of the lowest unoccupied and of the highest occupied molecular orbital, i.e. E<sub>LUMO</sub> and E<sub>HOMO</sub>, respectively, together with their difference ( $\Delta E_{L-H}$ ). E<sub>LUMO</sub> and E<sub>HOMO</sub> measure the ability of a molecule to accept and to donate an electron pair, respectively; thus, they describe the electrophilicity and the nucleophilicity of the substrate [109]. The difference  $\Delta E_{L-H}$  is a stability index: the higher  $\Delta E_{L-H}$ , the higher the compound reactivity in chemical reactions. In fact, it is the relative difference between the nucleophile and electrophile orbitals that governs the reactivity of a given nucleophile-electrophile interaction [110]. E<sub>HOMO</sub> appeared only in the QSAR of Log V<sub>max</sub> for FMO, with a positive coefficient, as expected from its catalytic cycle. The substrates of FMO are nucleophiles, i.e. electron donors [58], and the higher the HOMO energy, the greater is the ability of the chemical to act as an electron donor. E<sub>LUMO</sub> and  $\Delta E_{L-H}$  featured in QSARs both for Log (1/K<sub>m</sub>), generally with a negative correlation coefficient and for Log V<sub>max</sub>, with a positive correlation. This could be explained with the kinetics of the Michaelis-Menten reactions. Both K<sub>m</sub> and V<sub>max</sub> can be expressed in terms of k<sub>cat</sub>: V<sub>max</sub> is the product of k<sub>cat</sub> and total enzyme concentration, and K<sub>m</sub> is equal to the ratio (k<sub>cat</sub> + k<sub>-1</sub>)/k<sub>1</sub>, where k<sub>-1</sub> and k<sub>1</sub> are constants, respectively, for breakdown and formation of the complex enzyme-substrate [24]. The more reactive the molecule (i.e. the higher  $\Delta E_{L-H}$ ), the higher is the catalytic rate (k<sub>cat</sub>), therefore the lower 1/K<sub>m</sub> (negative coefficient) and the higher V<sub>max</sub> (positive coefficient). The presence of E<sub>LUMO</sub> in the QSARs for Log V<sub>max</sub> for ADH, ALDH and CYP indicates that their substrates are likely electrophilic in nature, as it can be expected from their metabolic reactions. For ADH, a network of hydrogen bonding interactions facilitates the deprotonation of the alcohol substrate bound to the active site of the enzyme [111]. The ALDH catalytic mechanism involves a nucleophilic attack on the carbonyl group (C=O) of the aldehydes [112], which are reactive electrophilic compounds. At the CYP active site, the oxidation of chemicals is carried out by an electron-deficient complex (FeO<sub>3</sub><sup>+</sup>), which abstracts either a hydrogen atom or an electron from the substrate [49]. CYP enzymes would then behave as Lewis bases (nucleophiles) or Brønsted

bases (H-acceptors). In fact, together with  $E_{\text{LUMO}}$ , also  $\text{pK}_{\text{a}}$  and hydrogen bonding properties were important in the QSARs for  $\text{Log } V_{\text{max}}$  in CYP.

#### 4.5 Conclusions

The QSARs developed in this study for  $\text{Log } (1/K_{\text{m}})$  and  $\text{Log } V_{\text{max}}$  of four important oxidising enzymes included physicochemical descriptors, which can be calculated and interpreted in a straightforward way. The processes underlying biotransformation were discussed from a mechanistic point of view, which may be useful in future research aimed at the prediction of the clearance of chemicals.

#### Appendices

Appendix B contains the datasets collected for this study, as well the formulas of the statistical parameters used to assess model fitting and predictivity.

Appendix D contains the non-standardised regression coefficients and the applicability domains of the QSAR models.

#### Acknowledgements

Alessandra Pirovano was supported by the European Union through the Environmental ChemOinformatics (ECO) Project (FP7-PEOPLE-ITN-2008, no. 238701). Karin Veltman was supported by the European Union through a Marie-Curie Intra-European Fellowship (FP7-PEOPLE-2010-IEF, no. 273104). This research was partly funded by the European Union through the TOX-TRAIN Project (implementation of a TOXicity assessment Tool for pRActical evaluation of life-cycle Impacts of techNologies) (FP7-PEOPLE-2011-IAPP, no. 285286).

## Chapter 5

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

Alessandra Pirovano

Stefan Brandmaier

Mark A.J. Huijbregts

Ad M.J. Ragas

Karin Veltman

A. Jan Hendriks

*Environmental Toxicology and Pharmacology* (2015), 39(1), 247-258

## 5.1 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 [1, 113]. 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. Firstly, the chemical needs to reach the enzyme and bind to it; secondly, a catalytic reaction must occur. The latter process is described by the maximum rate of reaction ( $V_{\max}$ ) at saturating substrate concentration [25]. 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 ( $CL_{\text{int}}$ ). 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 [19, 114].

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 [115]. In previous studies, metabolic constants were frequently found to correlate with easily interpretable physicochemical properties of substrates, such as hydrophobicity or hydrogen bonding [116]. However, the reported QSARs had generally low explained variances [117] or considered only a limited series of substrates [26]. Weak correlations indicated that the metabolic processes could only partly be explained by the physicochemical descriptors chosen, possibly because of the complexity of the underlying metabolic reactions [102]. 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 [118]. 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.

## 5.2 Materials and Methods

### 5.2.1 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 [105]. 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 [58]. P450 enzymes usually catalyse monooxygenase reactions, which involve the insertion of an oxygen atom into a substrate [60].

Data were taken from the BRENDA enzyme database [57] and several reviews [26, 58-60]. 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 Appendix B (Table B1).

$K_m$  values were expressed in  $\mu\text{M}$  and all rates were expressed as  $V_{max}$  with  $\mu\text{mol min}^{-1} \text{mg}_{\text{PROT}}^{-1}$  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 B2 in the Appendix B).

$K_m$  and  $V_{max}$  data were combined into 4 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.

### 5.2.2 Molecular descriptors

Approximately 2000 descriptors were calculated using the Online CHEmical Modeling environment platform (OCHEM) [91]. These descriptors included Mopac descriptors (version 7.1) [119], E-state indices (electro-topological state indices) [120], ALogPS [121], Adriana code (<http://www.molecular-networks.com>), Chemaxon (<http://www.chemaxon.com>), CDK [122], Spectrophores (Silicos NV, <http://openbabel.org>) and a subset of Dragon 6 (constitutional, topological and information indices, geometrical, charge, 3D-MorSE and GETAWAY descriptors, 2D autocorrelations, functional group counts, atom-centred fragments, molecular properties) [123].

### 5.2.3 Model development and validation

First, the data of each dataset were split into a training set and a validation set in a 2:1 proportion [124]. For each training set, we calculated the correlation coefficient ( $R$ ) of each descriptor with the experimental  $\text{Log}(1/K_m)$  and  $\text{Log } V_{max}$  values and filtered out descriptors with  $|R| < 0.4$ . This procedure assures the stability and reliability of the models because only descriptors that have some correlation with the endpoint are considered.

A Genetic Algorithm (GA) was then applied to the remaining descriptors with WEKA v.3.6.7 [125] to find the optimal subsets of variables that yielded models with the highest predictive powers [126]. The following parameters were set for the GA: 30 cycles, 400 children and 75 survivors. Because of the relatively small number of compounds in the datasets and to avoid overfitting, the number of variables to select was prefixed and limited to a maximum of 6 for ALDH, FMO and CYP, or four for ADH. The GA was optimised on multiple linear regression (MLR) and included a leave-one-out (LOO) validation procedure. With LOO cross validation, a single observation is removed from the original dataset, and the remaining observations are used as training data such that each observation is removed only once. A model is then developed for each reduced data set, and the response values of the removed observations are predicted from these models. The fitness function of the GA was the correlation coefficient for the LOO validation ( $Q_{LOO}$ ): for each of the datasets, the subsets of one to six variables that provided the highest  $Q_{LOO}$  were selected.

The Akaike's information criterion (AIC) was calculated to select which of the models with one to six variables was the most adequate to predict the  $\text{Log } V_{\text{max}}$  and the  $\text{Log } (1/K_m)$  of each enzyme. The AIC is a trade-off between a good fit to the model (measured by the likelihood) and a penalty for complexity (calculated using the number of parameters). The model with the lowest AIC is interpreted as the best model. The collinearity of the descriptors was checked using variance inflation factors (VIFs) calculated with the R package 'car' [127]. The threshold for collinearity was  $\text{VIF} > 5$  [128]. Therefore, for each dataset we selected the model with the lowest AIC and having all variables with  $\text{VIFs} < 5$ .

The models were first developed using the original values of the descriptors to obtain regression coefficients that can be used to estimate the  $K_m$  and  $V_{\text{max}}$  values for other chemicals. However, the descriptors are expressed in different units and scales, therefore those coefficients do not indicate the importance of each model parameter. To determine this importance, the predictors were scaled to zero-mean and unit-variance (auto-scaling) and used to calculate the standardised regression coefficients of the models. The values of the standardised coefficients allow for comparison of the contribution of each descriptor in influencing  $K_m$  and  $V_{\text{max}}$ . In addition, the predictors were classified into four general categories: 1) Functional group or fragment (E-state, functional group counts, etc.); 2) Size and shape (topological and geometrical descriptors); 3) Partitioning ( $\log P$ ); or 4) electronic parameters (descriptors related to electronic properties such as charge, polarizability, etc.).

For every model, the coefficient of determination ( $R^2$ ) and the Root Mean Squared Error (RMSE) were calculated as measures of model fit. The applicability domains of the QSARs, required by the OECD QSAR validation principles [64], are defined by the range (min and max) of the values of the descriptors used to build the model [104].

The MLR models developed using the training sets were validated with the WEKA data mining software using two procedures: leave-one-out (LOO) cross validation and external cross validation with the validation set. The predictive ability of the models was quantified using the  $R^2$  and the RMSE for the LOO cross-validation ( $Q^2_{\text{LOO}}$  and  $\text{RMSE}_{\text{LOO}}$ ) and for the external validation ( $R^2_{\text{EXT}}$  and  $\text{RMSE}_{\text{EXT}}$ ). The equations used to calculate the statistical parameters are reported in Appendix B.

### 5.3 Results

For every enzyme, the QSAR models selected for  $\text{Log } (1/K_m)$  and  $\text{Log } V_{\text{max}}$  and their statistical parameters are provided in Tables 5.1 and 5.3, respectively. Tables 5.2 and 5.4 contain the definitions of the descriptors used in the QSARs

and their categories with brief explanations when necessary. In the equations of the QSARs, the variables are reported in order of relative importance from highest to lowest. Figure 5.1 shows the values of the standardised regression coefficients of the predictors selected for A)  $\text{Log } (1/K_m)$  and B)  $\text{Log } V_{\max}$ . Figures 5.2 and 5.3 compare the measured values to the values predicted by the QSARs for  $\text{Log } (1/K_m)$  and  $\text{Log } V_{\max}$ , respectively. The applicability domains of the QSARs are provided in Tables E1 and E2 of Appendix E for  $\text{Log } (1/K_m)$  and  $\text{Log } V_{\max}$ , respectively.

### 5.3.1 $\text{Log } (1/K_m)$

The models for  $\text{Log } (1/K_m)$  had explained variances ( $R^2_{\text{adj}}$ ) and leave-one-out cross-validated explained variances ( $Q^2_{\text{LOO}}$ ) of approximately 50% for CYP and FMO, 70% for ALDH and 80% for ADH (Table 5.1). The predictive abilities of the models ( $R^2_{\text{ext}}$ ) were approximately 50% for CYP and FMO and 60% for ADH and ALDH (Table 5.1). For ADH, the number of aliphatic secondary alcohols (nOHs, Dragon 6) was the most important descriptor (i.e. the one with the highest standardised correlation coefficient, negative in this case). The most influential descriptor for ALDH was the Adriana 3D autocorrelation descriptor 3DACorr\_PiChg\_2 with a positive coefficient. For FMO, the most important descriptor was RHSA, a CDK descriptor combining surface area and partial charge information, which was positively correlated with  $\text{Log } (1/K_m)$ . For CYP, the most important descriptor was the aromaticity index AROM (Dragon 6) with a negative coefficient.

### 5.3.2 $\text{Log } V_{\max}$

The best models for  $\text{Log } V_{\max}$  had explained variances ( $R^2_{\text{adj}}$ ) and leave-one-out cross-validated explained variances ( $Q^2_{\text{LOO}}$ ) varying from approximately 20% for FMO to approximately 80% for ADH (Table 5.3). The explained variances were approximately 50% and 60% for ALDH and CYP, respectively. The predictive abilities of the models ( $R^2_{\text{ext}}$ ) were approximately 30% for FMO, 50% for CYP and ALDH and 60% for ADH (Table 5.3). For ADH and ALDH, the most important descriptors were the functional group counts nHDon and nArX (Dragon 6), respectively, which were both negatively correlated with  $\text{Log } V_{\max}$ . These descriptors indicate the number of donor atoms for hydrogen bonds (nHDon) and the number of halogens on an aromatic ring (nArX). For FMO and CYP, the most influential descriptors were the E-state indices Se1C3N3as and Se1C1C3sd, with a positive and a negative coefficient, respectively.



Figure 5.1. Standardised regression coefficients of the predictors in the QSARs for (A) Log (1/K<sub>m</sub>) and (B) Log V<sub>max</sub> for the four enzyme classes (ADH, ALDH, FMO and CYP). The standardised coefficients were obtained by using the descriptors scaled to zero-mean and unit-variance. The predictors were classified in four categories: 1) Functional group or fragment; 2) Size and shape; 3) Partitioning; 4) electronic property.

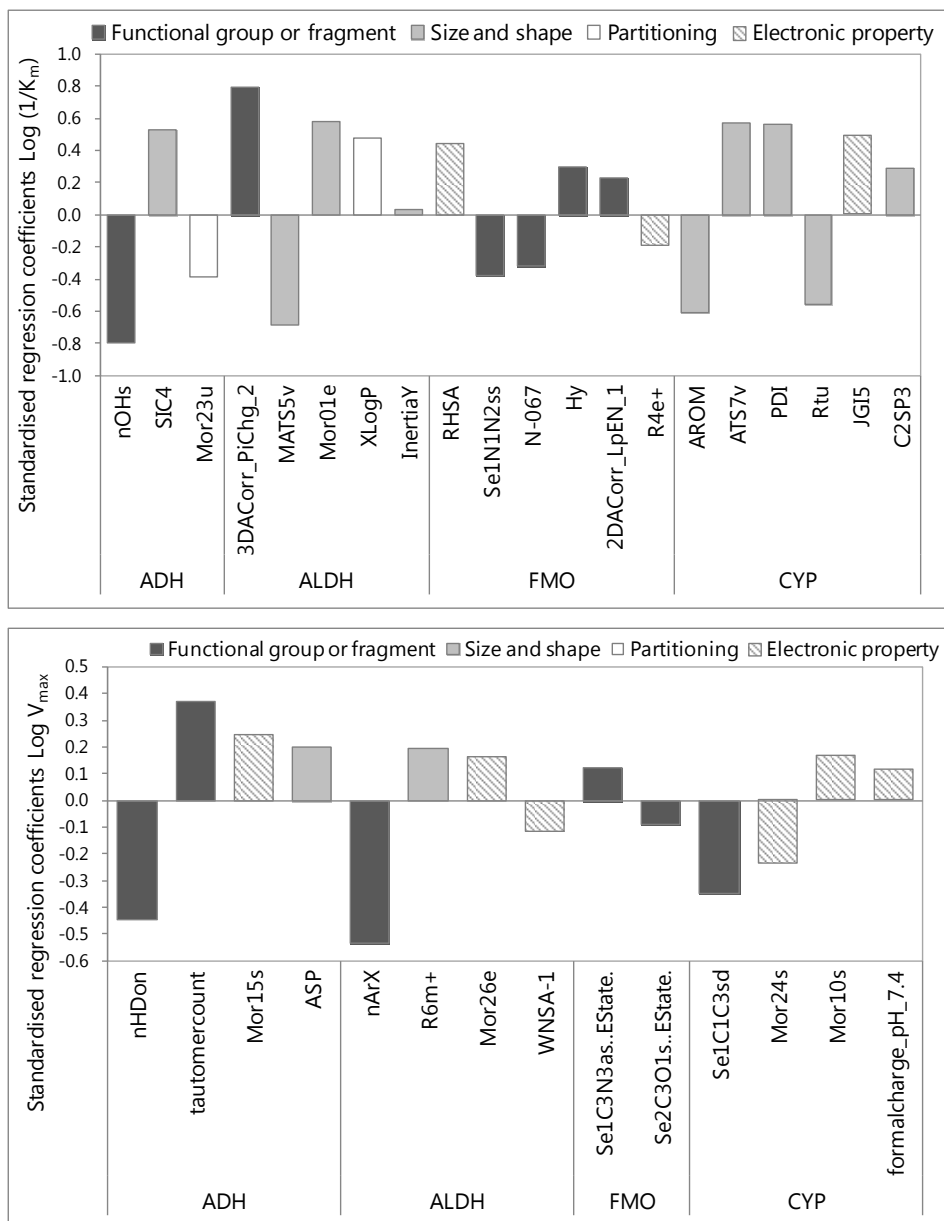


Table 5.1. Quantitative Structure Activity Relationships (QSARs) for Log (1/K<sub>m</sub>). The variables are reported in order of relative importance.

Name	QSAR	n <sub>TEST</sub>	R <sup>2</sup>	R <sup>2</sup> <sub>adj</sub>	RMSE	P	Q <sup>2</sup> <sub>LOO</sub>	RMSE <sub>LOO</sub>	n <sub>EXT</sub>	R <sup>2</sup> <sub>EXT</sub>	RMSE <sub>EXT</sub>
ADH_Km	-1.66(±0.23) nOHs +2.30(±0.48) <b>SIC4</b> -	24	0.85	<b>0.83</b>	0.47	<1E-7	<b>0.82</b>	0.53	10	<b>0.59</b>	0.77
	0.64(±0.19) <b>Mor23u</b> -4.35(±0.37)										
ALDH_Km	21.30(±3.79) <b>3DACorr_PiChg_2</b> -	52	0.73	<b>0.70</b>	0.84	<1E-11	<b>0.68</b>	0.93	25	<b>0.64</b>	0.94
	3.04(±0.60) <b>MAT55v</b> +5.0E-3(±1.3E-3)										
	<b>Mor01e</b> +0.32(±0.09) <b>XLogP</b> +7.9E-										
	5 <sup>o</sup> (±3.3E-4) <b>InertiaY</b> -1.60(±0.30)										
FMO_Km	4.05(±0.81) <b>RHSA</b> -0.26(±0.06) <b>Se1N1N2ss</b>	99	0.51	<b>0.48</b>	0.68	<1E-11	<b>0.45</b>	0.73	50	<b>0.54</b>	0.67
	-0.90(±0.23) <b>N-067</b> +0.29(±0.11) <b>Hy</b> +5.4E-										
	3(±2.0E-3) <b>2DACorr_LpEN_1</b> -5.80(±2.46)										
	<b>R4e+</b> -5.37(±0.84)										
CYP_Km	-1.62(±0.40) <b>AROM</b> +0.64(±0.17) <b>ATS7v</b>	81	0.56	<b>0.52</b>	0.59	<1E-10	<b>0.50</b>	0.63	40	<b>0.47</b>	0.63
	+8.32(±1.85) <b>PDI</b> -0.14(±0.04) <b>RTu</b>										
	+17.48(±3.24) <b>JGI5</b> +0.16 <sup>o</sup> (±0.08) <b>C2SP3</b> -										
	8.71(±1.38)										

<sup>a</sup> The probability (p) value of the coefficient is greater than 0.05.

Table 5.2. Explanation of the descriptors in the QSARs for Log (1/K<sub>m</sub>)

Enzyme	Name	Group	Definition	Classification
ADH	nOHs	Dragon 6 (Functional groups)	Number of secondary alcohols (aliphatic)	Functional group or fragment
	SIC4	Dragon 6 (Information indices)	Structural Information Content index (neighbourhood symmetry of 4-order)	Size and shape. It is a topological index encoding information on the 2D structure.
	Mor23u	Dragon 6 (3D-MoRSE)	3D-MoRSE - signal 23 / unweighted	Partitioning. It is negatively correlated with logP (Dragon 6) (R<0.9) for the compounds in the training set.
ALDH	3DACorr_PiChg_2	Adriana (Spatial or 3D property-weighted autocorrelation descriptors)	3D autocorrelation weighted by $\pi$ atom charges.	Functional group or fragment. It is positively correlated (R>0.85), among others, to nArNO2 (Dragon 6), which is the number of nitrogen groups in an aromatic molecule.
		Dragon 6 (2D autocorrelations)	Moran autocorrelation of lag 5 weighted by van der Waals volume	Size and shape. It describes how a certain property (in this case van der Waals volume, representing the shape) is distributed along the topological structure (2D).
	Mor01e	Dragon 6 (3D-MoRSE)	signal 01 / weighted by Sanderson electronegativity	Size and shape. It is positively correlated with the molecular surface area (Chemaxon) (R>0.95).
	XLogP	CDK	Octanol-water partitioning coefficient predicted by the XLogP atom-type method	Partitioning

Continuation of Table 5.2

	InertiaY	Adriana (Shape and size descriptors)	Principal moment of inertia of second principal axis [Da·Å <sup>2</sup> ]	Size and shape
FMO	RHSA	CDK (Electronic and geometric descriptors)	Relative sum of solvent accessible surface areas of atoms with absolute value of partial charges less than 0.2	Electronic property
	Se1N1N2ss	E-state	Molecular Bond E-state index	Functional group or fragment. Single bond (e1) between 2 N atoms (N2 and N1), i.e. (R)-NH-NH2
	N-067	Dragon 6 (Atom-centred fragments)	Al2-NH	Functional group or fragment
	Hy	Dragon 6 (Molecular properties)	Hydrophilic factor	Functional group or fragment. It is highly correlated (R>0.9) to H-050 (Dragon 6), which is an atom-centred fragment related to H atoms attached to heteroatoms.
	2DACorr_LpEN_1	Adriana	2D autocorrelation weighted by lone pair electronegativities	Functional group or fragment. It is highly correlated (R>0.9) to nHet (Dragon 6), which is the number of heteroatoms.
	R4e+	Dragon 6 (GETAWAY descriptors)	R maximal autocorrelation of lag 4 / weighted by Sanderson electronegativity	Electronic property. It incorporates information on the 3D structure and weight the molecule atoms by Sanderson electronegativities (electronic).

Continuation of Table 5.2

Enzyme	Name	Group	Definition	Classification
CYP	AROM	Dragon 6 (Geometric descriptors)	Aromaticity index	Size and shape. It is a geometrical descriptor encoding information on the 3D structure.
	ATS7v	Dragon 6 (2D autocorrelations)	Broto-Moreau autocorrelation of lag 7 (log function) weighted by van der Waals volume	Size and shape. It describes how a certain property (in this case van der Waals volume, representing the shape) is distributed along the topological structure (2D).
	PDI	Dragon 6 (Molecular properties)	Packing Density Index	Size and shape. It is the ratio between the McGowan volume and the total surface area.
	RTu	Dragon 6 (GETAWAY Descriptors)	R total index / unweighted	Size and shape. It encodes information on the 3D molecular structure.
	JGI5	Dragon 6 (2D autocorrelations)	Mean topological charge index of order 5	Electronic property
	C2SP3	CDK (Topological Descriptors)	Singly bound carbon bound to two other carbons	Size and shape. It is a topological index encoding information on the 2D structure.

Figure 5.2 (next page). Measured versus predicted  $\text{Log}(1/K_m)$  values in mammals for compounds metabolised by (A) ADH; (B) ALDH; (C) FMO; (D) CYP. The solid lines indicate the 1:1 bisector and the dashed lines indicate  $\pm 2$  log units error. Laboratory measurements (dots): Log transformed geometrical mean of  $1/K_m$  [ $\mu\text{M}^{-1}$ ] for each compound, with the geometric standard error (horizontal bar). The white dots represent the group of chemicals used in the external validation set.

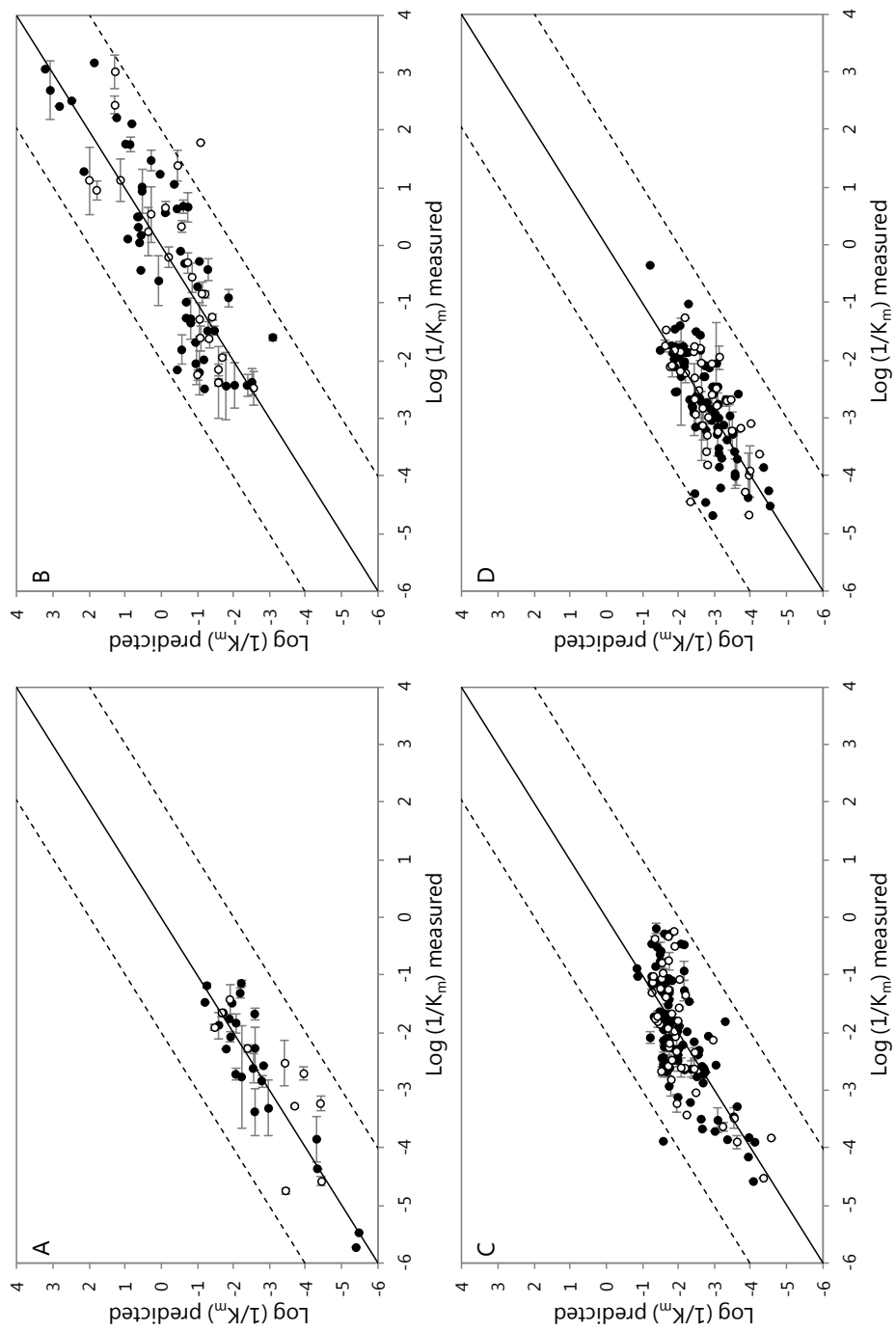


Table 5.3. Quantitative Structure Activity Relationships (QSARs) for Log  $V_{\max}$ . The variables are reported in order of relative importance.

Name	QSAR	n	R <sup>2</sup>	R <sup>2</sup> <sub>adj</sub>	RMSE	p	Q <sup>2</sup> <sub>LOO</sub>	RMSE <sub>LOO</sub>	n <sub>EXT</sub>	R <sup>2</sup> <sub>EXT</sub>	RMSE <sub>EXT</sub>
ADH_V	-0.49(±0.08) <b>nHDon</b> +0.94(±0.15)	22	0.86	<b>0.82</b>	0.22	<1E-6	<b>0.75</b>	0.30	11	<b>0.65</b>	0.51
	<b>tautomercount</b> +0.14(±0.04) <b>Mor15s</b> +0.86(±0.26) <b>ASP</b> -0.82(±0.21)										
ALDH_V	-1.74(±0.26) <b>nArX</b> +3.56(±1.55) <b>R6m+</b> +1.27(±0.53) <b>Mor26e</b> -3.5E-3 <sup>a</sup> (±2.1E-3)	49	0.57	<b>0.53</b>	0.43	<1E-6	<b>0.50</b>	0.47	25	<b>0.48</b>	0.36
	<b>WNSA-1</b> -0.16 <sup>a</sup> (±0.20)										
FMO_V	0.07(±0.02) <b>Se1C3N3as</b> -0.04(±0.02)	61	0.24	<b>0.21</b>	0.28	<1E-3	<b>0.16</b>	0.30	31	<b>0.27</b>	0.28
	<b>Se2C3O1s</b> -0.21(±0.04)										
CYP_V	-0.55(±0.08) <b>Se1C1C3sd</b> -0.38(±0.07)	81	0.65	<b>0.63</b>	0.38	<1E-15	<b>0.62</b>	0.40	40	<b>0.48</b>	0.47
	<b>Mor24s</b> +0.13(±0.03) <b>Mor10s</b> +0.58(±0.24) <b>formalcharge_pH_7.4</b> -1.29(±0.05)										

<sup>a</sup> The probability (p) value of the coefficient is greater than 0.05.

Table 5.4. Explanation of the descriptors in the QSARs for Log  $V_{\max}$ 

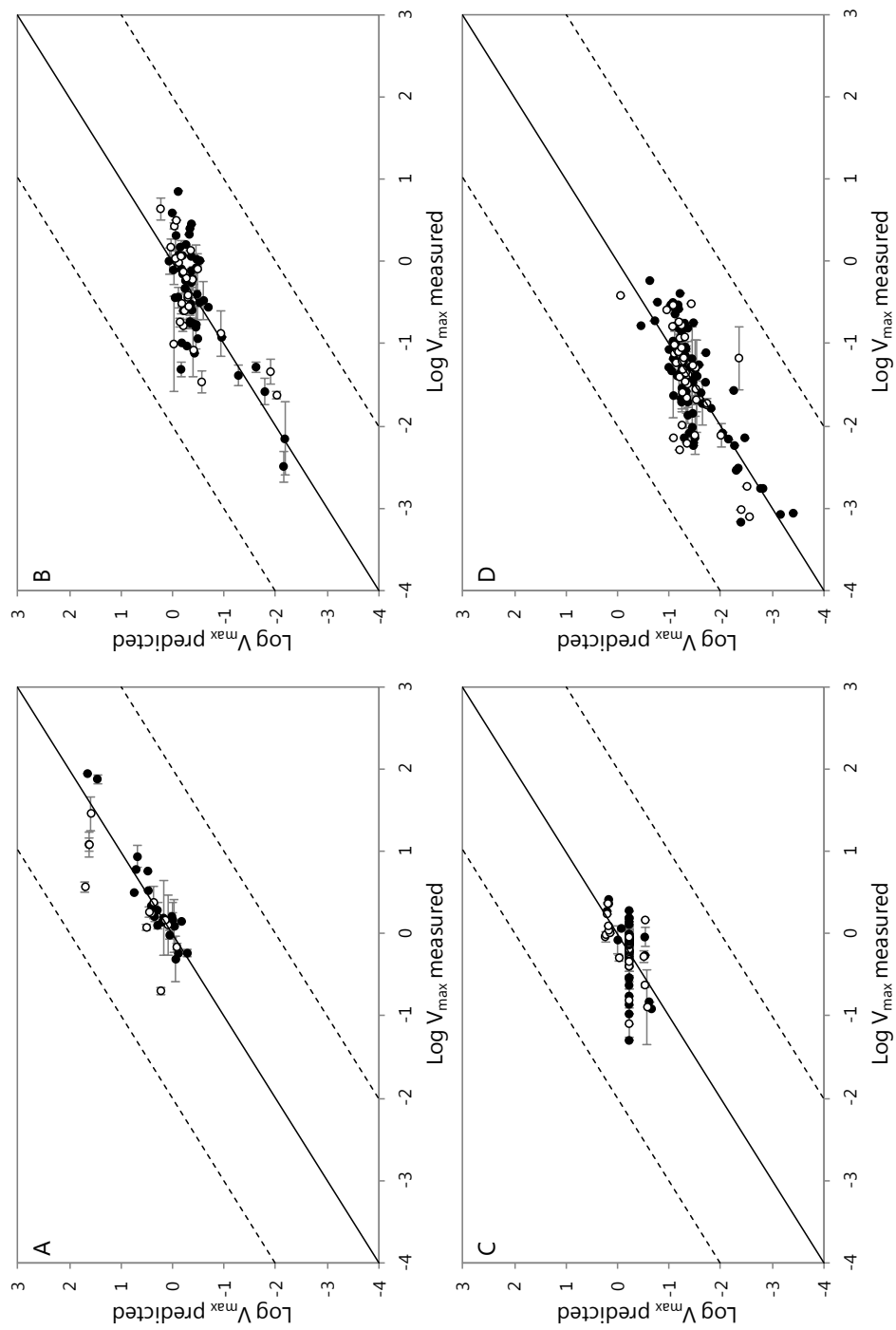
Enzyme	Name	Group	Definition	Property class
ADH	nHDon	Dragon 6 (Functional group counts)	Number of donor atoms for hydrogen-bonds (with N and O)	Functional group or fragment
	tautomercount	Chemaxon (Isomers)	Number of tautomers	Functional group or fragment. It is highly correlated ( $R > 0.9$ ) to Se1C2C2sd, an E-state index representing a single bond (e1) between an sp2 C and an sp3 C ( $X = C-C-R$ ).
	Mor15s	Dragon 6 (3D-MoRSE descriptors)	Signal 15 / weighted by I-state	Electronic property. It incorporates information on the 3D structure and weights the molecule atoms by ionisation state (electronic).
	ASP	Dragon 6 (Geometrical descriptors)	Asphericity	Size and shape
ALDH	nArX	Dragon 6 (Functional group counts)	Number of halogens (X) on aromatic ring (Ar). In this case, $X = Cl, Br, I, F$	Functional group or fragment
	R6m+	Dragon 6 (GETAWAY descriptors)	R maximal autocorrelation of lag 6 / weighted by atomic masses	Size and shape. It incorporates information on the 3D structure and weights the molecule atoms by their masses (size).
	Mor26e	Dragon 6 (3D-MoRSE descriptors)	Signal 26 / weighted by atomic Sanderson electronegativities	Electronic property. It incorporates information on the 3D structure and weights the molecule atoms by Sanderson electronegativities (electronic).
	WNSA-1	CDK (Electronic and geometric descriptors)	Partial negative surface area weighted by total molecular surface area	Electronic property



Continuation of Table 5.4

FMO	Se1C3N3as	E-State	Molecular Bond E-state index	Functional group or fragment. Single bond (e2) between C (C3) and N (N3), with N attached to an aromatic ring.
	Se2C3O1s	E-state	Molecular Bond E-state index	Functional group or fragment. Double bond (e2) between C (C3) and O (O1), i.e. indicating a carbonyl group (R'(R)C=O).
CYP	Se1C1C3sd	E-state	Molecular Bond E-state index	Functional group or fragment. Single bond (e1) between 2 C atoms (C3 and C1), i.e. RC(=X)-C.
	Mor24s	Dragon 6 (3D-MoRSE descriptors)	Signal 24 / weighted by I-state	Electronic property. It incorporates information on the 3D structure and weights the molecule atoms by ionisation state (electronic).
	Mor10s	Dragon 6 (3D-MoRSE descriptors)	Signal 10 / weighted by I-state	Electronic property. It incorporates information on the 3D structure and weights the molecule atoms by ionisation state (electronic).
	formalcharge_pH_7.4	Chemaxon (Charge)	Formal charge of the molecule calculated at pH 7.4	Electronic property

Figure 5.3 (next page). Measured versus predicted Log  $V_{\max}$  values in mammals for compounds metabolised by (A) ADH; (B) ALDH; (C) FMO; (D) CYP. The solid lines indicate the 1:1 bisector and the dashed lines indicate  $\pm 2$  log units error. Laboratory measurements (dots): Log transformed geometrical mean of  $V_{\max}$  [ $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}_{\text{PROT}}^{-1}$ ] for each compound, with the geometric standard deviation (horizontal bar). The white dots represent the group of chemicals used in the external validation set.



## 5.4 Discussion

### 5.4.1 Model limitations

Because the experimental  $K_m$  values and rates were collected from the scientific literature, they come from different laboratories often employing different protocols, e.g. conditions of pH and temperature, which can affect enzyme activity [78]. In addition, the rates were reported in the papers either as  $V_{max}$  or  $k_{cat}$  values. The latter were transformed into  $V_{max}$  (Appendix B, Table B2) using the weight of the enzyme or the content of microsomal protein (for CYP) as conversion factors. For the conversion factors, we used the values reported in the studies measuring  $k_{cat}$ , when available; otherwise, we used the average values from other studies. Consequently, part of the residual error is likely caused by these different sources of variation in the input data (i.e. experimental variation and inaccuracies in conversions). Furthermore, we merged data measured for different mammalian species (i.e. human, horse, rat, mouse, pig and rabbit) and isoenzymes (i.e. any of the several forms of an enzyme, all of which catalyse the same reaction but are characterised by different properties). The merging process is likely another source of unexplained variation. Finally, when using a QSAR to predict the  $K_m$  or  $V_{max}$  value of a new compound, it is important to know whether the chemical is a putative substrate for the enzyme.

The QSARs developed for CYP in the present work yielded lower  $R^2$  values than the QSARs obtained in other studies with  $R^2$  values of approximately 0.8-0.9 [26, 27]. However, the latter datasets typically included homologous series of approximately 10 structurally related compounds metabolised by one given isoenzyme in one mammalian species. Thus, those models are only applicable to specific combinations of compounds, isoenzymes and species for which a similar behaviour can be anticipated. Cronin et al. [67] argued that an  $R^2$  value between 0.6 and 0.7 is all that can realistically be expected for heterogeneous datasets such as the ones used in the present study.

For a model with good external predictability,  $R^2_{ext}$  values should be higher than 0.5, and the difference between  $R^2$  and  $R^2_{ext}$  should be no larger than 0.2-0.3 [129]. This result was the case for all models except for the  $V_{max}$  of FMO. The low explained variance for the  $V_{max}$  of FMO is likely because of an unusual feature of its catalytic cycle, in which substrate binding has no effect on velocity [58]. The rate-limiting step of the FMO catalytic cycle depends on one of two initial enzyme reactions, i.e. either the reaction of the FAD prosthetic group with NADPH or its successive reaction with molecular oxygen. These two steps generate the enzyme-bound flavin-hydroperoxide (FADOOH) that is required before binding and responsible for the oxidation of suitable nucleophiles that gain access to the FMO catalytic site. Because the rate-

limiting step for the overall reaction rate occurs before substrate oxidation,  $V_{\max}$  is independent of chemical properties. Consequently, the  $V_{\max}$  values of FMO are generally similar across different chemicals, whereas the  $K_m$  values may vary [130]. In this study,  $V_{\max}$  values covered less than two orders of magnitude ( $-1.3 < \text{Log } V_{\max} < 0.4$ , Fig. 5.3C and Table E2 in the Appendix E).

#### 5.4.2 Model interpretation

##### *Log (1/ $K_m$ )*

The importance of the properties influencing  $1/K_m$  appeared to be specific to the enzyme group considered. Functional groups or fragments were the most relevant predictors for the enzyme groups metabolising specific compounds, i.e. ADH, ALDH and FMO, which have substrates that are mainly alcohols, aldehydes and chemicals with a nucleophilic heteroatom, respectively. These predictors provide information on the chemical features that drive substrate binding. For ADH, the most influential descriptor nOHs (Dragon 6) indicates the number of aliphatic secondary alcohols (R-CH-OH-R) that are metabolised into ketones by ADH. The binding affinity is lower for secondary alcohols, as shown by the negative regression coefficient of nOHs, possibly because the OH group on the secondary carbon disfavours the hydrophobic interaction between the alkyl groups of the substrates and the active site of ADH enzymes. For ALDH, the most important descriptor 3DACorr\_PiChg\_2 (Adriana) was positively correlated ( $R > 0.85$ ) with the number of nitrogen groups in an aromatic molecule (nArNO<sub>2</sub>, Dragon 6). Log (1/ $K_m$ ) values are higher for aromatic aldehydes (positive regression coefficient), which are usually also more hydrophobic. Functional groups or fragments were particularly relevant for Log (1/ $K_m$ ) of FMO (four of the six selected descriptors). The E-state index Se1N1N2ss and the Dragon 6 descriptor N-067 refer to nitro groups. These fragments represent single bonds between two N atoms (NH<sub>2</sub>-NH) and the number of fragments containing secondary aliphatic amines, respectively. FMO substrates are typically soft nucleophiles, i.e. compounds with functional groups bearing a polarizable, electron-rich centre that is usually a heteroatom (such as nitrogen, sulphur and phosphorus) in organic compounds [58]. The descriptor 2DACorr\_LpEN\_1 (Adriana) was highly related ( $R > 0.9$ ) to the number of heteroatoms (nHet, Dragon 6). The positive coefficient shows that the higher the number of heteroatoms in the molecule, the higher the chances for the substrate to bind to FMO are, consistent with FMO catalytic cycle. The hydrophilic factor (Hy, Dragon 6) describes the hydrogen-bond donor ability of the molecules. This predictor is related to the presence of hydrophilic groups in the molecule, which comprise hydrogen attached to an electronegative heteroatom (-OH, -SH, -NH). The  $1/K_m$  increases with the hydrogen-bond donor

ability, suggesting the importance of hydrogen bonding in the interactions of the molecule with the binding site of the enzyme.

In all QSARs for  $\text{Log}(1/K_m)$  except for FMO, the majority of the predictors were associated with partitioning or the size and shape of the substrates. These descriptors indicate the importance of weak, non-specific interactions between substrate and binding site of these enzymes, e.g. via desolvation processes, consistent with previous work [30, 79]. In particular, for CYP enzymes, four of five predictors were related to the geometry of the molecules, likely because of the broad substrate specificity of these enzymes, which can bind to and oxidise many structurally diverse compounds. In fact, any electron-donating substrate that is properly positioned can gain access to the CYP active site [131]. In the  $\text{Log}(1/K_m)$  QSAR for CYP, the predictors AROM (Dragon 6) and C2SP3 (CDK) represent the aromaticity index and the number of single bond carbon atoms bound to two other carbon atoms, respectively. Aromatic molecules comprise planar rings of  $\text{sp}^2$  hybridized atoms with a cyclic electron delocalisation that makes these compounds stable [132]. These two predictors are related to the number of aromatic atoms, which describes a hydrophobic feature of the molecules and was positively correlated with  $1/K_m$ . The packing density index (PDI) is a molecular property defined as the ratio between the McGowan volume and the total surface area. The positive correlation coefficient of PDI shows that binding increases with substrate size for CYP enzymes. The molecular surface area featured in the QSAR for ALDH (Mor01, Dragon 6) was positively correlated with  $1/K_m$ . A larger molecular size increases the possibility of interactions with the binding site and the hydrophobic nature of the molecules. The descriptors for partitioning are Mor23u (Dragon 6) for ADH and XLogP for ALDH. The latter is the octanol-water partitioning coefficient ( $\log P$ ) predicted using the XLogP atom-type method (CDK) and had a positive regression coefficient. Mor23u (Dragon 6) was negatively correlated with  $\log P$  (Dragon 6) for the compounds in the training set ( $R < -0.9$ ). For ADH and ALDH,  $1/K_m$  increased with increasing hydrophobicity of the substrates, confirming the hydrophobic nature of the binding site of these enzymes.

The electronic parameters were relevant for FMO and, to a lesser extent, for CYP. This result indicates that for these enzymes  $1/K_m$  describes strong interactions with substrates, such as polar bonds, which can be understood from their catalytic cycles [117]. The catalytic mechanism of FMO involves a nucleophilic attack that occurs before binding [58]. CYP enzymes have a catalytic mechanism with many steps occurring between binding and substrate oxygenation [49].  $K_m$  values may be sensitive to kinetic perturbations at catalytic steps occurring after substrate binding; thus,  $1/K_m$  values may not be good approximations of affinity constants [107]. For CYP, the 2D

autocorrelation descriptor JGI5 belongs to the Galvez topological charge indices, which evaluate the charge transfers between pairs of atoms and the global charge transfers in the molecule [133]. For FMO, the most important descriptor RHSA (CDK) is a combination of surface area and partial charge. RHSA was positively correlated with  $1/K_m$  as polar interactions increase with increasing solvent accessible area occupied by partial charges. Another descriptor of electronic properties for FMO was R4e+, which is a GETAWAY Dragon descriptor weighted by Sanderson electronegativity (e). This result confirms the importance of partial charges in the interaction of substrates with FMO enzymes.

### *Log V<sub>max</sub>*

The “functional groups or fragments” descriptors were particularly important for Log V<sub>max</sub>. For Log V<sub>max</sub> of ADH, the most influential descriptors nHDon (Dragon 6) and tautomercount (Chemaxon) indicate the number of donor atoms for hydrogen bonds (i.e. the count of N and O atoms) and the number of tautomers (i.e. isomers that have the same molecular formula but switching single bond and adjacent double bond), respectively. The V<sub>max</sub> tends to be higher for chemicals with a lower hydrogen-bond donor ability, which correspond to the aldehydes in the ADH dataset. The number of tautomers is also a fragment that is linked to aldehydes. This descriptor was highly correlated ( $R>0.9$ ) with Se1C2C2sd, an E-state index representing a single bond (e1) between an sp<sup>2</sup> C and an sp<sup>3</sup> C ( $X=C-C-R$ ). For the compounds in the dataset, this bond was found in the aldehyde fragments ( $R-C=O$ ), and its positive regression coefficient again indicates that aldehydes yield higher V<sub>max</sub> values. In fact, ADH enzymes metabolise also aldehydes to alcohols at a rate that is higher than the one of the opposite reaction (from alcohols to aldehydes). For ALDH, the most important predictor nArX (Dragon 6) represents the number of halogens ( $X = Cl, Br, I, F$ ) on an aromatic ring. The compounds that have this fragment are characterised by a low Log V<sub>max</sub>, as shown by the negative regression coefficient of nArX. Notably, these compounds are halogenated benzaldehydes, compounds that were outliers for the Log ( $1/K_m$ ) regressions with Log K<sub>ow</sub> in our previous work [79]. This result can be expected because compounds containing more halogens (particularly Cl and F) are usually more stable. For CYP, the most important predictor Se1C1C3sd describes a single bond between two C atoms ( $RC(=X)-C$ ). This descriptor was highly correlated ( $R>0.9$ ) with the Dragon 6 descriptor H-051, which represents the number of H atoms attached to alpha C (i.e. the C atom bonded to a functional group). The negative sign of the regression coefficient shows that a lower number of H atoms attached to alpha C increases the velocity of the reaction. The alpha C is an active atom and tends to lose acidic protons, thus affecting the reactivity of the substrates [134]. For FMO, all

descriptors selected for  $\text{Log } V_{\text{max}}$  were E-state indices. Se1C3N3as indicates tertiary amines (N3). The presence of this fragment increases the  $V_{\text{max}}$  (positive regression coefficient), suggesting that the nucleophilic attack is favoured on this nitrogen. Se2C3O1s represents the carbonyl group ( $\text{R}'(\text{R})\text{C}=\text{O}$ ); in the FMO dataset, this fragment recurs in amides ( $\text{R}'\text{-NC(=O)-R}$ ) and carbamate pesticides ( $\text{R}'\text{-NC(=O)O-R}$ ). The negative regression coefficient suggests that the presence of the carbonyl group in the molecule lowers its maximum velocity. Because the statistics for the  $\text{Log } V_{\text{max}}$  QSAR for FMO were not satisfactory, these descriptors can be considered only an indication of the involvement of nitrogen in substrates metabolised by FMO.

Electronic properties of the substrates also played an important role in the QSARs for  $\text{Log } V_{\text{max}}$ . Interactions characterised by the cleavage and formation of covalent or ionic bonds are described by electronic properties of the substrates. For all enzymes except FMO, at least one Dragon 6 3D-MoRSE descriptor (3D-Molecule Representation of Structures based on Electron diffraction) was selected. MoRSE descriptors yield good modelling power for different biological and physicochemical properties because they simultaneously consider the 3D structure and various atomic properties [135]. Polarity was relevant for the  $V_{\text{max}}$  of CYP substrates, as indicated by the descriptor `formalcharge_pH_7.4` (Chemaxon). Previous studies on P450 enzymes have also demonstrated that  $V_{\text{max}}$  depends on electronic properties [29]. At the CYP active site, the oxidation of chemicals is performed by an electron-deficient complex ( $\text{FeO}_3^+$ ), which abstracts either a hydrogen atom or an electron from the substrate [49]. Therefore, strong interactions are involved in the maximum velocity of these enzymes.

For  $\text{Log } V_{\text{max}}$ , only a few descriptors related to the size and shape of the molecules were featured in the QSARs. Furthermore, their occurrences were limited to the models for ADH and ALDH. For ADH, ASP (molecular asphericity) describes the shape of molecules; it varies from zero for totally spherical molecules to unity for flat molecules, such as benzene. The positive regression coefficient shows that flat molecules are characterised by higher values of  $\text{Log } V_{\text{max}}$ , likely because more reactive sites are accessible to the metabolising enzymes. For ALDH, the geometry predictor R6m+ (Dragon 6) belongs to GETAWAY (GEometry, Topology and Atom-Weights Assembly) descriptors weighted by atomic mass. These descriptors are based on spatial autocorrelation formulae that incorporate 3D information and weight the molecule atoms by different properties, such as mass, polarizability and volume [135]. The small role played by geometric factors in determining  $V_{\text{max}}$  compared with fragments and electronic properties is because of the nature of enzymatic catalysis. Metabolic reactions are characterised by bond cleavage and formation, which are better explained by electronic factors. In addition,

functional groups or fragments can capture the features of the substrates that are involved in the chemical- and enzyme-specific mechanisms of metabolic reaction.

## 5.5 Conclusions

The importance of the properties influencing the affinity constant ( $1/K_m$ ) appeared to be specific to the enzyme group considered. Functional groups or fragments were the most relevant predictors for the enzyme groups metabolising specific compounds, i.e. ADH, ALDH and FMO. Size and shape properties were also important for binding, especially for CYP enzymes, likely because of the broad substrate specificity of CYP enzymes. These descriptors indicate weak non-specific interactions between the substrates and binding sites of these enzymes, e.g. via desolvation processes. Electronic factors and functional groups or fragments were particularly important for the maximum reaction rate  $V_{max}$ . This constant represents the catalytic process, which involves specific interactions between substrate and enzyme, characterised by the cleavage and formation of covalent bonds. 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.

## Appendices

Appendix B contains the datasets collected for this study, as well the formulas of the statistical parameters used to assess model fitting and predictivity.

Appendix E contains the applicability domains of the QSAR models.

## Acknowledgements

Alessandra Pirovano was supported by the European Union through the Environmental ChemOinformatics (ECO) Project (FP7-PEOPLE-ITN-2008, no. 238701). Karin Veltman was supported by the European Union through a Marie-Curie Intra-European Fellowship (FP7-PEOPLE-2010-IEF, no. 273104). This research was partly funded by the European Union through the TOX-TRAIN Project (implementation of a TOXicity assessment Tool for pRActical evaluation of life-cycle Impacts of techNologies) (FP7-PEOPLE-2011-IAPP, no. 285286).



## **Chapter 6**

# **QSARs for estimating intrinsic hepatic clearance of organic chemicals in humans**

Alessandra Pirovano

Stefan Brandmaier

Mark A.J. Huijbregts

Ad M.J. Ragas

Karin Veltman

A. Jan Hendriks

Submitted

## 6.1 Introduction

Biotransformation is one of the processes that can influence the bioaccumulation of compounds in organisms [136]. Through biotransformation, the parent compound is converted via enzymatic reactions into another chemical (metabolite), which is usually more soluble and thus can be excreted more easily [1]. The biotransformation potential of xenobiotics is often assessed using data from *in vitro* metabolic tests [19, 137, 138]. Since liver is the principal organ responsible for metabolism in fish and mammals, *in vitro* assays are mostly performed with preparations from hepatic tissue, such as isolated hepatocytes, S9 liver fractions, or liver microsomes [22, 139, 140]. The xenobiotics are incubated with these liver preparations, which contain different complements of metabolising enzymes, to obtain the *in vitro* intrinsic clearance ( $CL_{INT}$ ). For reactions that exhibit classical Michaelis-Menten kinetics and at non-saturating substrate concentrations, the *in vitro*  $CL_{INT}$  is defined as the ratio between the maximum velocity of the reaction ( $V_{max}$ ) and the Michaelis constant ( $K_m$ ), which is the substrate concentration at half  $V_{max}$  [19]. The *in vitro*  $CL_{INT}$  values can be extrapolated to estimate whole-body *in vivo* biotransformation rates, thus they can be of crucial importance for the risk assessment of xenobiotics [19, 138].

Measured *in vitro*  $CL_{INT}$  data are available only for a limited number of chemicals and species, and models can be useful to predict the  $CL_{INT}$  for chemicals that have not been tested yet. Quantitative structure-activity relationships (QSARs) are models correlating structural, physical and chemical properties of substances with their biological activity by means of statistical approaches [10]. QSARs are based on the assumption that compounds with similar structural features will have similar biological activities and/or physicochemical properties. The models built on experimental data can then be used to predict the biological activity of a broader range of related chemicals. Other advantages of QSARs, beyond prediction, include identifying influential structural and/or physicochemical characteristics or gaining insights into the mechanism of action for the process investigated [10]. Models have been built to predict enzyme-specific  $K_m$  and  $V_{max}$  of various xenobiotics metabolised by oxidising enzymes in mammals [141]. These QSARs are important to predict and to understand the enzymatic processes underlying specific metabolic pathways. It is, however, often difficult to know beforehand which metabolic pathway(s) a substance will undergo. For this reason, clearance measured in liver preparations containing different complements of enzymes, such as microsomes and hepatocytes, provide a more accurate measurement of the overall metabolic activity. QSARs have been developed to predict clearance in microsomes or hepatocytes of mammals [33-36] using information on the chemical structure. Nevertheless, these models included

only pharmaceuticals, with the aim to accelerate the selection of new candidates in the drug discovery stage based on their predicted clearance. To our knowledge, no QSARs have yet been developed to predict *in vitro* CL<sub>INT</sub> including environmental pollutants in the training set.

The aim of this study was to develop QSARs for *in vitro* clearance in humans measured in hepatocytes and microsomes. The QSAR models were based on datasets of 118 compounds (of which 53 environmental pollutants) for hepatocytes and 115 compounds (of which 56 environmental pollutants) for microsomes. The models were built with multiple linear regressions (MLR) by selecting theoretical descriptors and were mechanistically interpreted to provide insight into the processes governing biotransformation. External validation was applied to assess the predictive power of the models [64].

## 6.2 Materials and Methods

### 6.2.1 Experimental dataset

Clearance data (CL<sub>INT</sub>) for humans were collected from the scientific literature for the two most commonly used *in vitro* metabolism assays: isolated hepatocytes and liver microsomes [142]. Liver microsomes are subcellular fractions (endoplasmatic reticulum) with relatively high concentrations of phase I drug-metabolising enzymes, especially cytochrome P450 (CYP) [143]. Isolated hepatocytes are liver cells, thus they contain the full complement of phase I and phase II metabolic enzymes and essential cofactors (e.g. NADPH). Phase I enzymes metabolise most of the xenobiotics, so microsomes are often used to assess metabolism as they are convenient to prepare for many species. Nevertheless, predictions of *in vivo* CL<sub>INT</sub> from hepatocytes data are usually more accurate than those from microsomal data [144], since all possible metabolic reactions can take place in hepatocytes and most transporter functions are preserved, mimicking the *in vivo* systems<sup>[143]</sup>. Clearance can be measured either by the decrease in the amount of the parent compound (substrate depletion) or by an increase in the metabolites (product formation) [22]. The first method allows for a more precise quantification of the clearance, but data obtained with both methods were used in the present study in order to obtain larger datasets.

For hepatocytes, the measured CL<sub>INT</sub> values were taken from Tonnelier et al. [145], who gathered human liver metabolism data for 94 chemicals, mainly pesticides and drugs. Additional data were taken from Sohlenius-Sternbeck et al. 2010 [146], who measured CL<sub>INT</sub> values for 52 pharmaceuticals in human hepatocytes. All CL<sub>INT</sub> data collected (units:  $\mu\text{L}/\text{min}/10^6$  cells) were derived following substrate depletion. Only CL<sub>INT</sub> data with a quantified value were

retained, i.e. different from zero and above the limit of the detection. When more than one  $CL_{INT}$  value was available for one compound, the geometric mean of the  $CL_{INT}$  values was used in the dataset. The  $CL_{INT}$  values for human hepatocytes are reported in Table F1 (Appendix F), for a total of 119 compounds.

For liver microsomes, the measured  $CL_{INT}$  values were collected from individual studies published in scientific literature. We used the following search terms in the Pubchem and Google Scholar search engines (last access on 7 November 2014): 1) liver, human *in vitro*, microsomes, and 2) intrinsic clearance,  $V_{max}$ ,  $K_m$ , Michaelis-Menten, first-order, rate constant, kinetic constant, kinetic rate. We checked all papers resulting from this search (approximately 6,000), together with all of the citing and cited papers. Among these papers, we retained only those reporting experiments conducted in human liver microsomes at physiological conditions, i.e. pH 7.4 and T 37°C. All data were expressed as  $CL_{INT}$  (units:  $\mu\text{L}/\text{min}/\text{mg}_{\text{MICR}}$ ); if data were reported as  $K_m$  and  $V_{max}$ ,  $CL_{INT}$  was calculated as the ratio  $V_{max}/K_m$ . The majority of the data for environmental pollutants (more than 90%) was measured following product formation, while for pharmaceuticals clearances were all determined following substrate depletion. For the experiments following product formation, if more than one main metabolite was detected, the clearance of the parent compound was calculated as the sum of the clearance values measured for each product. When more than one  $CL_{INT}$  value was available for one compound, the geometric mean of the  $CL_{INT}$  values was used in the dataset. The  $CL_{INT}$  values for human liver microsomes are reported in Table F2 (Appendix F), for a total of 115 compounds.

### 6.2.2 Molecular descriptors

The datasets were uploaded to the Online CHEmical Modeling environment platform [91] (OCHEM, <http://ochem.eu>) and the chemical structures were visualised to check if they were correct. In addition, the nitro groups on the molecules were standardised to  $\text{N}(=\text{O})=\text{O}$  [147, 148]. Approximately 2200 descriptors were calculated using the OCHEM platform, including

1. E-state indices [149], which combine electronic and topological information about a molecule and allow identification of the relevant structural fragments governing the activity of chemicals.
2. The octanol/water partition coefficient (LogP) and solubility in water (LogS) with the ALOGPS 2.1 program [121].
3. Chemaxon descriptors at pH 7.4 [150], including elemental analysis (e.g. mass, atom count, etc.), charge, geometry (e.g. polar surface area, volume, etc.), partitioning (i.e. LogD7.4), acceptor and donor counts, etc.

4. MOPAC descriptors [119] (version 7.1): MOPAC is a semi-empirical molecular orbital package which allows the calculation of quantum chemical descriptors such as HOMO and LUMO energies, electronic energy, etc.
5. DRAGON descriptors [132] (version 6), including only constitutional (mass, atom and bond counts, etc.), topological and geometrical descriptors, connectivity indices, functional group counts, atom-centred fragments, charge descriptors and molecular properties.
6. Adriana code (<http://www.molecular-networks.com>), including physicochemical, 2D, 3D and surface-based molecular descriptors and properties.
7. CDK [122], including constitutional (atom and bond counts) and topological descriptors.

One substance (abamectin) was omitted from the hepatocytes dataset because not all molecular descriptors could be calculated. This was due to the fact that the CDK package was unable to process this big molecule.

#### *6.2.3 Model development and validation*

The QSAR models were developed following the same steps both for hepatocytes (118 compounds) and microsomes data (115 compounds). Before developing the QSARs, the  $CL_{INT}$  value of each chemical was Log transformed in order to normalise the data [35]. For each dataset, the data were split into a training set and a test set in a 2:1 proportion [124]. Chemicals were ordered according to decreasing values of clearance and separated into triplets. From each of the triplets, one chemical was inserted in the test set (33% of the compounds): for the hepatocytes, it was the second compound of each triplet and for the microsomes it was the third one. The compounds in each training set were used to build the QSAR, which was applied to the compounds in the test set to estimate the predictive power of the model.

For each training set, ten descriptors were selected. A common forward selection was implemented for the prioritisation of the most relevant combination of descriptors with the p-value (derived from a general linear regression model) as the decisive criterion whether to include the descriptor. General linear models (GLM) were developed with the software R v.3.03 [95]. The R package 'bestglm' [96] was used to select the best subset among the 10 descriptors after an exhaustive search. In order to avoid overfitting, the maximum number of variables to be included in the subsets was set at 6 [97]. The collinearity of the variables was checked using variance inflation factors (VIFs), calculated with the R package 'car' [127]. If all variables had  $VIFs < 5$  [128], the QSAR was accepted, which was always the case.

The models were first developed using the original values of the descriptors to obtain regression coefficients that can be used to estimate the *in vitro* CL<sub>INT</sub> values for other chemicals. However, the descriptors are expressed in different units and scales, therefore the resulting coefficients do not indicate the importance of each model parameter. To determine this importance, the predictors were scaled to zero-mean and unit-variance (auto-scaling) and used to calculate the standardised regression coefficients of the models. The values of the standardised coefficients allow for comparison of the contribution of each descriptor in influencing CL<sub>INT</sub>. In order to facilitate the interpretation of the models, the predictors were classified into four general categories: (1) functional group or fragment (E-state, functional group counts, etc.); (2) size and shape (topological and geometrical descriptors); (3) partitioning (Log P, LogD<sub>7.4</sub>); or (4) electronic parameters (descriptors related to electronic properties such as charge, polarizability, etc.).

The fitting ability of the QSARs was evaluated using a range of statistical parameters, i.e. the coefficient of determination ( $R^2$ ), the adjusted  $R^2$  ( $R^2_{adj}$ ), the Root Mean Squared Error (RMSE) and the p-value from the F-test (p). The applicability domains of the QSARs, required by the QSAR validation principles established by the Organisation for Economic Co-operation and Development (OECD) [64], were defined by the range (min and max) of the values of the descriptors used to build the model [104]. The models built using the training sets were validated with WEKA using two procedures: internal validation of the models with the leave-one-out (LOO) procedure and external cross-validation of the models with the test set. The LOO cross-validated  $R^2$  ( $Q^2_{LOO}$ ) and RMSE (RMSE<sub>LOO</sub>) were calculated to assess the internal predictivity of the models and the external predictivity was expressed with the external coefficient of determination ( $R^2_{ext}$ ) and RMSE (RMSE<sub>ext</sub>). The equations used to calculate the statistical parameters are reported in Appendix B.

### 6.3 Results

The resulting QSAR models for hepatocytes and microsomes are reported in Table 6.1, together with their statistical parameters. Table 6.2 contains the definitions of the descriptors used in the QSARs and their categories. In the equations of the QSARs, the variables are reported in order of relative importance from highest to lowest standardised regression coefficients. Figure 6.1 shows the values of the standardised regression coefficients of the predictors selected for CL<sub>INT</sub>. Figure 6.2 compares the measured Log CL<sub>INT</sub> values to the values predicted by the QSARs for A) human hepatocytes and B) human microsomes. The applicability domains of the QSARs are provided in Table F3 of Appendix F.

Significant correlations ( $p < 0.01$ ) were obtained for both the hepatocytes and microsomes QSARs (Table 6.1), with explained variances ( $R^2_{adj}$ ) of 67% and 50%, respectively. The leave-one-out cross-validated explained variances ( $Q^2_{LOO}$ ) and the predictive abilities of the models ( $R^2_{ext}$ ) were approximately 60% for hepatocytes and 30% for microsomes (Table 6.1). The most important variables were R5e+ for hepatocytes and HATS5e for microsomes, both with a negative regression coefficient. These are Dragon 6 GETAWAY descriptors weighted by Sanderson electronegativity, thus related to electronic properties. The other descriptors selected for hepatocytes were the Dragon6 GETAWAY HATS0m and R8u+, associated to fragments and geometry respectively, and the Adriana 2D autocorrelation descriptors 2DACorr\_SigChg\_2 and 2DACorr\_SigChg\_5 weighted by  $\sigma$  atom charges, thus related to electronic properties. The other descriptors selected for microsomes were the E-state indices Se2C2O1s and Se2O1P4s associated to fragments, the Dragon6 GETAWAY descriptor GATS4v associated to size, the Adriana 2D autocorrelation descriptor 2DACorr\_SigChg\_9 weighted by  $\sigma$  atom charges and the Chemaxon geometry descriptor SmallestRingSize.

Figure 6.1. Standardised regression coefficients of the predictors in the QSARs for Log CL<sub>INT</sub> for human hepatocytes and microsomes. The standardised coefficients were obtained by using the descriptors scaled to zero-mean and unit-variance. The predictors were classified in four categories: (1) Functional group or fragment; (2) Size and shape; (3) Partitioning (no descriptors selected); (4) Electronic property.

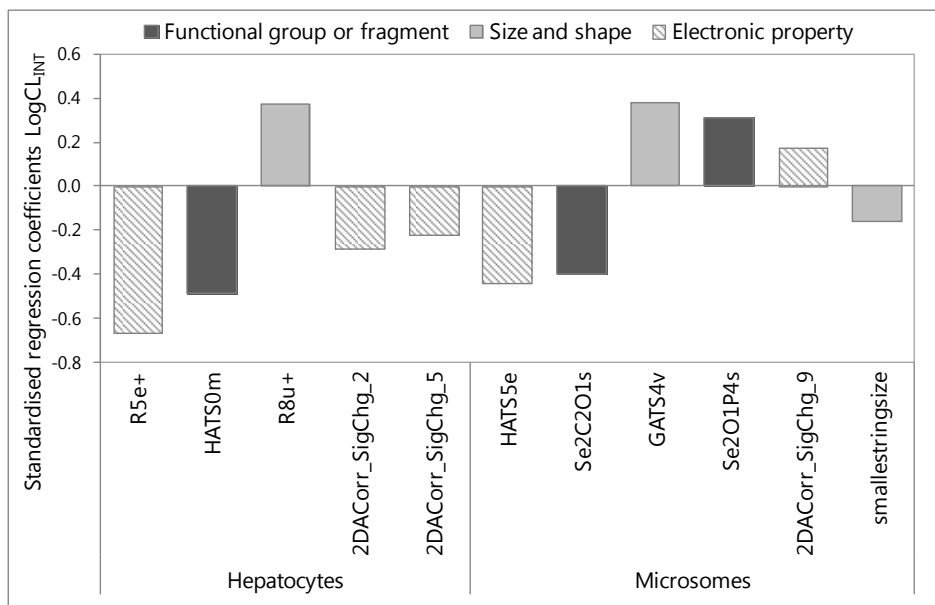


Table 6.1. QSARs for Log CL<sub>INT</sub>. The variables are reported in order of relative importance.

QSAR	n <sub>tr</sub>	R <sup>2</sup> <sub>adj</sub>	RMSE	p	Q <sup>2</sup> <sub>Loo</sub>	RMSE <sub>Loo</sub>	n <sub>ext</sub>	R <sup>2</sup> <sub>ext</sub>	RMSE <sub>ext</sub>
<b>HEPATOCYTES: -32.72 (±5.65) R5e+ -0.78(±0.19) HATS0m</b>									
+45.69(±10.74) R8u+ -1.39(±0.41) 2DACorr_SigChg_2 -1.07(±0.40)	79	<b>0.67</b>	0.68	<0.01	<b>0.62</b>	0.76	39	<b>0.62</b>	0.80
2DACorr_SigChg_5 +0.78(±0.33)									
<b>MICROSOMES: -1.71(±0.33) HATS5e -0.27(±0.05) Se2C2O1s</b>									
+1.32(±0.35) GATS4v +0.31(±0.07) Se2O1P4s +2.02(±0.86)	77	<b>0.50</b>	0.60	<0.01	<b>0.29</b>	0.76	38	<b>0.30</b>	0.70
2DACorr_SigChg_9 -0.07(±0.04) smalleststringsize+1.09(±0.28)									

Table 6.2. Explanation of the descriptors in the QSARs for Log CL<sub>INT</sub> for human hepatocytes and microsomes.

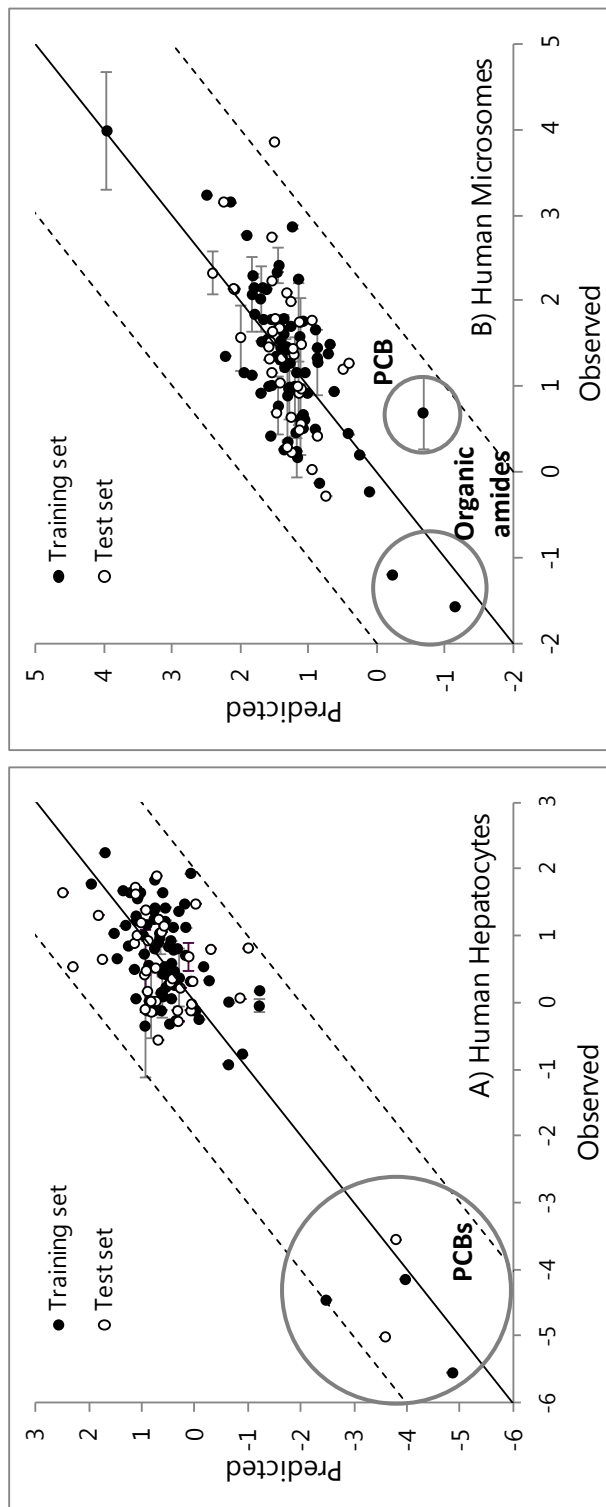
Name	Group	Definition	Classification
<b>HEPATOCYTES</b>			
R5e+	Dragon6 (GETAWAY descriptor)	R maximal autocorrelation of lag 5/ weighted by Sanderson electronegativity	Electronic property. It incorporates information on the 3D structure and weights the molecule atoms by Sanderson electronegativities (electronic).
HATS0m	Dragon6 (GETAWAY descriptor)	Leverage-weighted autocorrelation of lag 0/ weighted by mass	Functional group or fragment. It is highly correlated (R > 0.9) to Se1C3Cl1a, which is a molecular bond E-state index related to single bonds between a carbon atom in an aromatic ring and a chlorine atom.
R8u+	Dragon6 (GETAWAY descriptor)	R maximal autocorrelation of lag 8 / unweighted	Size and shape. It encodes information on the 3D molecular structure.
2DACorr_SigChg_2	Adriana (2D property-weighted autocorrelation)	2D autocorrelation of lag 2 weighted by $\sigma$ atom charges	Electronic property. Vectorial molecular descriptor derived from the 2D structure of a molecule and atom pair properties, in this case $\sigma$ atom charges (electronic).



Continuation of Table 6.2

2DACorr_SigChg_5	Adriana (2D property-weighted autocorrelation)	2D autocorrelation of lag 5 weighted by $\sigma$ atom charges	Electronic property. Vectorial molecular descriptor derived from the 2D structure of a molecule and atom pair properties, in this case $\sigma$ atom charges (electronic).
<b>MICROSOMES</b>			
HATS5e	Dragon6 (GETAWAY descriptor)	Leverage-weighted autocorrelation of lag 5/ weighted by Sanderson electronegativity.	Electronic property. It incorporates information on the 3D structure and weights the molecule atoms by Sanderson electronegativities (electronic).
Se2C2O1s	E-state	Molecular bond E-state index	Functional group or fragment. Double bond between an oxygen atom and a carbon atom bound to a substituent group and to an hydrogen atom ( $O=CH(R1)$ ).
GATS4v	Dragon 6 (2D autocorrelation)	Geary autocorrelation of lag 4 weighted by van der Waals volume.	Size and shape. It describes how a certain property (in this case van der Waals volume, representing the shape) is distributed along the topological structure (2D)
Se2O1P4s	E-state	Molecular bond E-state index	Functional group or fragment. Double bond between an oxygen atom and a pentavalent phosphorous atom.
2DACorr_SigChg_9	Adriana (2D property-weighted autocorrelation)	2D autocorrelation of lag 9 weighted by $\sigma$ atom charges	Electronic property. Vectorial molecular descriptor derived from the 2D structure of a molecule and atom pair properties, in this case $\sigma$ atom charges (electronic).
SmallestRingSize	Chemaxon (Geometry)	Number of atoms in the smallest ring.	Size and shape.

Figure 6.2. Measured versus predicted  $\text{Log CL}_{\text{INT}}$  values in human for: A) hepatocytes; B) microsomes. Datasets divided between training set (filled dots) and test set (white dots). Clearance expressed as  $\mu\text{L}\cdot\text{min}^{-1}\cdot 10^6\text{cells}^{-1}$  for hepatocytes and  $\mu\text{L}\cdot\text{min}^{-1}\cdot\text{mg}_{\text{MICR}}^{-1}$  for microsomes. Laboratory measurements (dots): Log transformed geometrical mean of  $\text{CL}_{\text{INT}}$  for each compound, with standard error (horizontal bar). Solid lines indicate the 1:1 bisector and dashed lines indicate  $\pm 2$  Log units error.



## 6.4 Discussion

### 6.4.1 Model limitations

In this study, *in vitro* clearance data measured in human hepatocytes and microsomes for pharmaceuticals and environmental chemicals were collected from literature. We used the data collected to build QSARs to predict *in vitro* clearance for a broader set of related chemicals using theoretical molecular descriptors. To our knowledge, this is the first attempt to predict this endpoint for such a diverse set of chemicals. The QSAR models were validated for predictivity (both internal and external) and an applicability domain was provided (Table F3 in Appendix F).

Because the experimental  $CL_{INT}$  values and rates were collected from individual papers, they come from different laboratories often employing different protocols and this can affect enzyme activity [78]. In addition, the rates were measured following substrate depletion in some cases and product formation in others. The QSARs developed for  $CL_{INT}$  in the present work yielded lower explained variances than the QSARs obtained in other models for hepatocytes [34-36], listed in Table 6.3, with  $R^2$  of approximately 0.8-0.9 and  $R^2_{ext}$  of 0.7-0.8. However, the previous QSARs were built with small datasets of 18 up to 71 pharmaceuticals and either included a large amount of descriptors compared to a small number of compounds in the training set (potential over-fitting), as was the case for [35] and [36], or used  $CL_{INT}$  values measured under standardised laboratory conditions, as was the case for the 18 compounds in [34]. Cronin et al. [67] argued that an  $R^2$  value between 0.6 and 0.7 is all that can realistically be expected for heterogeneous datasets such as the ones used in the present study. For a model with good external predictability,  $R^2_{ext}$  values should be higher than 0.5, and the difference between  $R^2$  and  $R^2_{ext}$  should be no larger than 0.2-0.3 [129]. This was the case for the hepatocytes model, whereas the microsomes QSAR had lower explained variance ( $R^2_{adj}$  50% vs. 67% and  $R^2_{ext}$  30% vs. 62%, Table 6.1). This may be because the data set for microsomes was more heterogeneous. For microsomes, data were obtained from different studies (almost one study per compound, implying a large experimental variability), while the data for hepatocytes are from standardised experiments (most of the 118 compounds were measured in two studies).

Table 6.3. Summary of QSARs models presented in literature to predict *in vitro* hepatocytes clearance using molecular descriptors (modified from [35]).

Year	Source	Statistical method	Descriptors in the models	Training (test)	Model performance
2000	[34]	MLR	4: electronic properties	18 (26)	$R^2 = 0.88$ ; RMSE = 0.28; $R^2_{\text{ext}} = 0.79$
2009	[35]	MLR	13 descriptors: molecular properties, constitutional, topological, geometrical descriptors, information indices, electrostatic properties	36 (13)	$R^2 = 0.85$ ; RMSE = 0.28, $R^2_{\text{ext}} = 0.73$
2010	[36]	ANN	21 descriptors: molecular properties, constitutional, topological, geometrical descriptors, information indices, WHIM descriptors	71 (18)	$R^2 = 0.91$ ; RMSE = 0.24, $R^2_{\text{ext}} = 0.65$

ANN = Artificial Neural Networks; MLR = Multiple Linear Regression

#### 6.4.2 Model interpretation

The intrinsic clearance in human liver microsomes and hepatocytes is a composite rate determined by various factors: chemical-specific uptake kinetics (either by transporters or by passive diffusion), chemical-specific association-dissociation kinetics with the metabolising enzymes, the actual chemical reaction rate and the 'free' or unbound chemical fraction available to interact with the enzymes [143]. In addition, the metabolic rate is influenced by the enzyme composition in the *in vitro* assay, i.e. both concentration of individual enzymes and which enzymes are present. In fact, chemicals can be metabolised by more than one enzyme, each with different specialities and reaction characteristics, and it is difficult to know beforehand which metabolic pathway they will undergo. Therefore, in the interpretation of the QSARs all these factors need to be considered.

In both QSARs, electronic properties of the substrates played a dominant role in predicting the clearance, while partitioning properties were absent (Figure 6.1). This may suggest that processes usually influenced by weak interactions (such as passive uptake for hepatocytes and enzyme binding) are not rate-limiting. Interactions characterised by the cleavage and formation of covalent or ionic bonds are described by electronic properties of the substrates. Thus, partial charges are important in the catalytic reaction between substrate and enzyme, as also noted in previous QSARs for clearance of drugs in hepatocytes (Table 6.3) [34-36]. All electronic descriptors, except 2DACorr\_SigChg\_9, are negatively related to metabolic clearance, i.e. an increased value of the descriptors will decrease the clearance. It is difficult to give a mechanistic explanation based on such composite, largely mathematical autocorrelation descriptors combining structural and electronic characteristics of the molecule, but some observations can be made. For example, R5e+ and HATS5e are autocorrelation descriptors of lag 5, with lag being the topological distance. This means that only those atoms that are exactly 5 path lengths separated are included to calculate the values for these descriptors. Larger molecules would typically have more of these atoms, whereas small molecules would have less or none of these atoms. So, larger molecules would likely have a higher score on R5e+ and HATS5e. These two descriptors are weighted by the electronegativity, and their negative regression coefficient indicates that a higher electronegativity would result in a higher clearance. Molecules having more atoms with a high electron density, thus more reactive centres, will probably have higher metabolic rates. In combination with the size characterisation described above, this suggests that small molecules with many partially charged atoms are more easily metabolised than large molecules with less reactive centres.

Functional groups or fragments were also relevant for the clearances in both hepatocytes and microsomes (Figure 6.1) and were useful to identify specific compounds having a deviating clearance compared to the others in the datasets. For hepatocytes, the GETAWAY Dragon 6 descriptor is highly correlated ( $R > 0.9$ ) to the E-state index Se1C3Cl1a indicating a single bond between a carbon atom in an aromatic ring and a chlorine atom. The presence of chlorine substituents in an aromatic ring lowers the clearance (negative regression coefficient), similarly to what happens for bacterial biodegradation. In fact, the resistance of chlorinated aromatic compounds to biodegradation generally increases with the degree of chlorination [151]. In the hepatocytes dataset, the E-state Se1C3Cl1a is indicative of polychlorinated biphenyls (PCBs), which have much lower clearance values than the other compounds (Figure 6.2A). In the microsome dataset, only one compound belongs to the PCB class which has the third lowest observed clearance value (Figure 6.2B).

PCBs are persistent pollutants, having high intrinsic elimination half-lives in humans of approximately 10–15 years [152]. For microsomes, two molecular bond E-state indexes were selected: Se2C2O1s and Se2O1P4s. The E-state index Se2C2O1s corresponds to a double bond between an oxygen and a carbon atom bound to a substituent group and to a hydrogen atom ( $\text{O}=\text{CH}(\text{R1})$ ). In our QSARs, compounds with this group are organic amides ( $\text{O}=\text{CH}-\text{N}(\text{R1})\text{R2}$ ) which have low clearance values (negative regression coefficient). Two of these compounds have the lowest observed clearance values in the microsome dataset (Figure 6.2B), and all three of them have the lowest predicted  $\text{Log CL}_{\text{INT}}$  values. The E-state index Se2O1P4s corresponds to a double bond between an oxygen and a pentavalent phosphorous atom ( $\text{O}=\text{P}\leq$ ). In our datasets, the only compound with this bond is the pesticide profenofos, which is a phosphorothiolate pesticide ( $\text{O}=\text{P}-\text{S}-\text{C}$ ) and has the highest clearance value among all the compounds metabolised by human microsomes (Figure 6.2B). This is not surprising as organophosphates are known to be metabolically instable, in fact they displaced persistent pesticides as DDT [153].

Few geometry descriptors featured in the QSARs for metabolic clearance (Figure 6.1). The Chemaxon descriptor SmallestRingSize selected for microsomes represents the number of atoms forming the smallest ring in the compound. This descriptor has a negative regression coefficient, which indicates that compounds with larger rings are less easily metabolised and compounds without rings have higher clearances. A lack of rings generally increases the flexibility of chemicals [132]. Linear chemicals may thus better adjust to the active site of the enzyme and be more easily metabolised. In previous QSARs for clearance of drugs in hepatocytes, the shape and size factors were not among the most influential descriptors [34–36], indicating a minor role of weak and non-specific interactions between substrate and enzymes. In our previous QSARs on  $K_m$  and  $V_{\text{max}}$  for different metabolising enzymes, size and shape factors were relevant only for  $K_m$  and less for the catalytic reaction  $V_{\text{max}}$  [141]. The small role played by geometric factors in determining  $\text{CL}_{\text{INT}}$  compared to electronic properties suggests that clearance rates are representing the catalytic rates, as already observed above from the absence of partitioning properties. Metabolic reactions are characterised by bond cleavage and formation, which are better explained by electronic factors.

#### 6.4.3 Practical application

The QSARs obtained in the present study can be helpful to predict the *in vitro*  $\text{CL}_{\text{INT}}$  values for human hepatocytes and liver microsomes. Information on hepatic clearance is essential for the extrapolation from *in vitro* to *in vivo* metabolism (ivive), useful for risk assessment. In order to express the clearances obtained from hepatocytes ( $\mu\text{L}/\text{min}/10^6$  cells) and microsomes

( $\mu\text{L}/\text{min}/\text{mg}_{\text{MICR}}$ ) in a common unit, the *in vitro*  $\text{CL}_{\text{INT}}$  values need to be multiplied by the *in vitro* system scaling factor (SF) to obtain the intrinsic clearance in the liver ( $\text{CL}_{\text{INT,liver}}$ ,  $\text{L}/\text{min}/\text{g}_{\text{LIV}}$ ). The *in vitro* SFs are hepatocellularity for hepatocytes (HP,  $10^6$  cells/ $\text{g}_{\text{LIV}}$ ) and protein concentration for microsomes (PL,  $\text{mg}_{\text{PROT}}/\text{g}_{\text{LIV}}$ ). For humans, SF values of  $99 \times 10^6$  cells/ $\text{g}_{\text{LIV}}$  for HP and  $32 \text{ mg}_{\text{PROT}}/\text{g}_{\text{LIV}}$  for PL have been estimated with a meta-analysis [154]. Then, liver  $\text{CL}_{\text{INT}}$  values should be multiplied by liver weight (LW,  $\text{g}_{\text{LIV}}/\text{kg}$ ), which for humans is on average  $25.7 \text{ g}_{\text{LIV}}/\text{kg}$  [155], to obtain the *in vivo* intrinsic clearance ( $\text{CL}_{\text{INT,vivo}}$ ,  $\text{L}/\text{min}/\text{kg}$ ). In order to be incorporated into mass balance bioaccumulation models, established physiologically based models can be further used to extrapolate the *in vitro* intrinsic clearance to whole body *in vivo* biotransformation rates ( $k_m$ ,  $\text{min}^{-1}$ ) [19].

While beyond the scope of the present study, a comparison of the magnitude of the clearance rates measured in hepatocytes and microsomes assays and the application of *in vivo* methods needs to be addressed in future studies. For microsomes, with 50% explained variance and 30% external predictivity, the QSAR can potentially be improved when more *in vitro* data become available from standardised experiments (possibly following substrate depletion). Despite these future efforts, the current study shows that the explained variance of 67% and external predictivity of 62% for hepatocytes is encouraging, allowing application of the outcomes in *in vitro* to *in vivo* extrapolation.

## Appendix

Appendix F contains the datasets collected for this study for human hepatocytes and microsomes, as well the applicability domains of the QSARs.

## Acknowledgements

Alessandra Pirovano was supported by the European Union through the Environmental ChemOinformatics (ECO) Project (FP7-PEOPLE-ITN-2008, no. 238701).





## Chapter 7

### Synthesis

## 7.1 Introduction

In environmental modelling, the prediction of the biotransformation rate is a difficult task due to the specific action of metabolism, which depends on the chemical and the enzyme involved and varies among individual organisms and species. The overall aim of this thesis was to develop QSARs for the prediction of biotransformation of xenobiotics in mammals based on their chemical properties. The relationships between metabolic activity and chemical structure were developed using different types of descriptors and for *in vitro* systems representing different levels of biological organization (isolated enzymes, hepatocytes and microsomes). The advantages and disadvantages of the QSAR descriptors and of the *in vitro* systems are presented in Section 7.2.

The *in vivo* biotransformation rate  $k_m$  of chemicals can be obtained using different methods, as discussed in Section 1.2.2. For example,  $k_m$  can be estimated as the difference between measured elimination rate constants and the sum of elimination rate constants predicted assuming no metabolism [20, 21]. Alternatively,  $k_m$  values can be estimated by extrapolating the metabolic constants measured *in vitro* to their whole-body *in vivo* equivalents using established physiologically based models [19]. In Section 7.3, an *in vitro-in vivo* extrapolation (ivive) scheme is first explained for tests with isolated hepatocytes and liver microsomes or isolated enzymes, which are the *in vitro* assays analysed in this thesis (Section 7.3.1). Second, this scheme is used to derive  $k_m$  values using the experimental clearance values collected for human microsomes and hepatocytes and the extrapolated  $k_m$  values were compared to *in vivo* measurements (Section 7.3.2). Finally, in Section 7.4 a method is discussed to quantify bioaccumulation potential of the metabolites without knowing their exact identity (i.e. molecular structure), based on the quantification of the change of hydrophobicity of the parent compound due to biotransformation (Chapter 2).

## 7.2 Tentative comparison of the QSARs

In this thesis, the relationships between metabolic activity and chemical structure were investigated using different types of descriptors: first  $K_{ow}$  only, then mechanistic descriptors and finally theoretical descriptors. These models were developed for systems representing different levels of biological organization (isolated enzymes, microsomes and hepatocytes). The advantages and disadvantages of the models developed in Chapters 3 to 6 are listed in Table 7.1, with regard to the different descriptors and the different *in vitro* assays considered.

Table 7.1. Advantages and disadvantages of the three different approaches used to derive metabolic constants.

QSAR	Pros	Cons
$K_m$ and $V_{max}$ from purified enzymes using mechanistic descriptors (Chapters 3 and 4)	<ul style="list-style-type: none"> <li>+ Mechanistic interpretation</li> <li>+ Insights into the affinity to single enzymes</li> <li>+ <math>K_m</math> and <math>V_{max}</math> can be easily derived for new chemicals, as values of predictors are widely available</li> </ul>	<ul style="list-style-type: none"> <li>- Relatively low explained variance</li> <li>- Not all metabolic pathways taken into account</li> <li>- Models were not validated, so predictions for new chemicals (which need to be putative substrates for the enzyme) could be unreliable</li> </ul>
$K_m$ and $V_{max}$ from purified enzymes using theoretical descriptors (Chapter 5)	<ul style="list-style-type: none"> <li>+ Better statistical results compared to mechanistic descriptors</li> <li>+ Insights into the affinity for enzymes and catalytic reactions</li> <li>+ Models were validated, so they can be used for predictive purposes</li> </ul>	<ul style="list-style-type: none"> <li>- Some descriptors difficult to interpret</li> <li>- Not all metabolic pathways taken into account</li> <li>- Most descriptors are difficult to calculate (commercial software) and chemicals need to be putative substrates for the enzyme</li> </ul>
$CL_{INT}$ from human hepatocytes and microsomes using theoretical descriptors (Chapter 6)	<ul style="list-style-type: none"> <li>+ Satisfying statistics for hepatocytes, so QSAR could be used for <i>in vivo</i></li> <li>+ <math>CL_H</math> values for the overall hepatic metabolism (for microsomes only P450), no need to know the metabolic pathway</li> <li>+ Models were validated, so they can be used for predictive purposes</li> <li>+ Inclusion of diverse environmental pollutants (previous studies in mammals were focused on pharmaceuticals only)</li> </ul>	<ul style="list-style-type: none"> <li>- Some descriptors difficult to interpret</li> <li>- Difficulty to differentiate all the factors influencing clearance (e.g. transport processes for hepatocytes, enzymatic reactions..)</li> <li>- Many descriptors are not widely available (e.g. commercial software)</li> <li>- Data available for relatively few compounds (about 100 for dataset): more experimental <i>in vitro</i> data are needed</li> </ul>

### 7.2.1 Advantages and disadvantages of the different descriptors

Models in this thesis were built using a “mechanistic” approach for  $K_m$  and  $V_{max}$  of different enzymes in mammals (Chapters 3 and 4), as well as using a “theoretical” approach for enzymatic  $K_m$  and  $V_{max}$  (Chapter 5) and for  $CL_{INT}$  from human hepatocytes and microsomes (Chapter 6). The main advantage of the first approach is that it enhances the understanding of the processes governing biotransformation, while the theoretical approach allows optimising the model performance for prediction (Table 7.1). Mechanistic descriptors are also widely available, while theoretical descriptors are often calculated with commercial software, thus they are not easily retrievable if they need to be calculate for new compounds.

The metabolic action consists of two steps: binding and catalytic reactions, represented by  $1/K_m$  and  $V_{max}$ , respectively. In addition, hepatocytes are liver cells; therefore, for these assays, also uptake (via passive diffusion or transporters) influences the clearance. Binding and partitioning processes take place through reversible or permanent bonding between the substance and enzyme active site or the cell membrane/transporters, in case of hepatocytes. Binding and passive diffusion usually involve weak interactions (e.g. van der Waals interactions or hydrogen bonding), except for substance binding to FMO (nucleophilic attack). On the contrary, catalytic reactions and active transport are governed by strong interactions (e.g. ionic bond or covalent bonding). Weak interactions are usually influenced by partitioning and size properties of the molecules, while strong interactions are governed by electronic factors [25, 65].

Based on the *a priori* knowledge of the mechanism of biotransformation, the metabolic constants were expected to be mainly influenced by the following properties:

- enzymatic  $K_m$ : partitioning properties and size, as well as electronic factors influencing binding;
- enzymatic  $V_{max}$ : electronic properties governing chemical reactivity;
- hepatocytes and microsomes clearance  $CL_{INT}$ : electronic properties, as well as partitioning and size, influencing clearance ( $CL_{INT} = V_{max}/K_m$ ) and uptake (for hepatocytes).

The regressions between  $1/K_m$  and hydrophobicity (Chapter 3) showed that binding increased with compound Log  $K_{ow}$ , which can be understood from the tendency to transform lipophilic compounds into more polar, thus more easily excretable metabolites. Mechanistic insight was provided by the analysis of the slopes. For most of the substrate classes of ADH, ALDH and CYP, the resulting slopes had 95% Confidence Intervals covering the value of 0.6, typically noted

in the regressions between protein-water distribution ( $\text{Log } K_{pw}$ ) and  $\text{Log } K_{ow}$ . A reduced slope (0.2-0.3) was found for FMO: this may be due to a different reaction mechanism involving a nucleophilic attack. When the relationships between  $1/K_m$  and more mechanistic descriptors (such as area, hydrogen bonding, etc.) were investigated (Chapter 4), partitioning and size properties were the most important properties influencing binding for ADH and ALDH. For CYP and FMO, electronic properties, together with size for CYP, played a greater role in influencing  $1/K_m$ , and this was explained in relation to the catalytic mechanism of the enzymes. For FMO, this might be because of the metabolic mechanism involving a nucleophilic attack. CYP enzymes have a catalytic mechanism with many steps occurring between substrate binding and oxygenation [49]. It was shown that  $K_m$  values may be sensitive to kinetic perturbations at catalytic steps taking place after substrate binding; thus,  $1/K_m$  values may not be good approximations of affinity constants [107]. In the relationships between  $V_{max}$  and mechanistic descriptors (Chapter 4), electronic properties such as dipole moment and LUMO energy were the most relevant. This can be explained by the nature of the catalysis, which is characterised by the cleavage and formation of covalent or ionic bonds, thus strong interactions.

While mechanistic descriptors were helpful to gain some insight into the processes governing biotransformation, the models had generally low explained variances ( $0.4 < R_{adj}^2 < 0.7$  for  $\text{Log } (1/K_m)$  and  $0.2 < R_{adj}^2 < 0.5$  for  $\text{Log } V_{max}$ ). This might indicate that the metabolic processes could only partly be explained by the physicochemical descriptors chosen, possibly because of the complexity of the underlying metabolic reactions [102]. The “theoretical” approach used to predict enzymatic  $K_m$  and  $V_{max}$  (Chapter 5) had better statistical performances ( $0.5 < R_{adj}^2 < 0.8$  for  $\text{Log } (1/K_m)$  and  $0.2 < R_{adj}^2 < 0.8$  for  $\text{Log } V_{max}$ ), but the interpretation of the descriptors selected was not straightforward, although some general interpretation of the QSARs was provided. 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  $V_{max}$  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. Besides the better statistics, an advantage of the models developed for enzymatic  $K_m$  and  $V_{max}$  is that external validation was performed, thus allowing extrapolation to other chemicals. In this case, it is however necessary to know whether the chemical is a putative substrate for

the enzyme, as well as whether it is within the applicability domain of the model.

The “theoretical” approach employed to build the QSAR for  $CL_{INT}$  of human hepatocytes and microsomes (Chapter 6) yielded satisfactory explained variances of 50% and 67%, respectively, but again the results were difficult to interpret. For both liver assays, clearance was predominantly determined by electronic properties, while size and shape were less important. As clearance is dependent on enzyme binding and membrane permeation (for hepatocytes), partitioning properties were expected to be influent in these QSARs, but they were not among the selected descriptors. The minor role of geometry and partitioning suggests that enzyme binding and, for hepatocytes, uptake across the membrane are not rate-limiting *in vitro*, thus clearance rates are representing the metabolic rate. Functional groups of fragments were useful to identify specific compounds that have a reaction rate significantly higher or lower compared to the other compounds, such as PCBs, which were poorly metabolised by hepatocytes and microsomes. The models were externally validated, thus they can be used to predict the *in vitro* hepatic clearance of other chemicals within the applicability domain.

In conclusion, “theoretical” approaches should be used to obtain models that are able to predict the metabolic constants of heterogeneous groups of chemicals, such as the ones analysed in this thesis. Nevertheless, a preliminary exploration using basic physicochemical parameters (such as Log  $K_{ow}$ , molecular size, etc.) as well as electronic features was helpful to explain the processes underlying biotransformation.

### 7.2.2 Advantages and disadvantages of the *in vitro* assays

In this thesis, QSARs were developed for systems representing different levels of biological organization (isolated enzymes, hepatocytes and microsomes). The  $K_m$  and  $V_{max}$  constants measured in enzymatic assays are a measurement of the metabolic potential relative to a specific pathway. The clearance values measured in microsomes and hepatocytes, instead, are related to the overall hepatic metabolism for hepatocytes and the first phase metabolism (mainly P450) for microsomes. As a consequence, when the QSAR for hepatocytes is used to predict the clearance of a new compound, it is not required to know its metabolic pathway. For microsomes, it should only be known whether Phase 1 is the dominant metabolic process. This is an advantage over the models for the enzymes, for which the chemical should be a putative substrate for the enzyme and this is often difficult to know (Table 7.1). In addition, *in vitro* clearance values for hepatocytes and microsomes can be extrapolated to *in vivo* clearance values that are comparable to the measured values. On the contrary, *in vivo* extrapolations performed using enzymatic constants might not

be reliable, as the *in vitro* assays contain higher concentration of isolated enzymes that do not reflect the *in vivo* situation. In Appendix G (Figure G1), a comparison between enzyme data and hepatocyte data collected for this thesis showed that the former are poor predictors of intrinsic liver clearances. The models developed for enzymes are however useful to have a better understanding of the processes taking place at the enzymatic level. It is also important to notice that the  $K_m$  and  $V_{max}$  data used to develop the QSARs for the different enzymes were averaged over different mammal species, while  $CL_{INT}$  were measured only in human hepatocytes and microsomes. The merging of data from different species is another source of variability, thus care should be taken when using these models to obtain enzymatic  $K_m$  and  $V_{max}$  values for a species for which no experimental data were available.

### 7.3 *In vitro* to *in vivo* extrapolation

Data from *in vitro* metabolic tests are used to determine biotransformation potential of drugs and environmental pollutants in mammals and fish [19, 138]. Since liver is the principal organ responsible for the metabolism [1], most *in vitro* systems are derived from hepatic tissue. The biotransformation potential is frequently assayed via the *in vitro* measurement of hepatic intrinsic clearance ( $CL_{INT}$ ) in isolated enzymes, microsomes, S9 fractions or hepatocytes [156]. Liver microsomes are subcellular fractions (endoplasmatic reticulum) with relatively high concentrations of Phase 1 drug-metabolising enzymes, especially cytochrome P450 (CYP). Liver S9 are subcellular fractions (microsomes and cytosol) containing cytosolic Phase 2 enzymes, such as glutathione S-transferase (GST). Isolated hepatocytes are liver cells, thus they contain the full complement of Phase 1 and Phase 2 metabolic enzymes. The rate of biotransformation of chemicals can be monitored either by the decrease in the amount of the substrate (parent compound) or by an increase in the products (metabolites) [22]. To be incorporated into mass balance bioaccumulation models, *in vitro*  $CL_{INT}$  values must be extrapolated to estimate *in vivo*  $k_m$  for the whole-body [19]. In this section, first a general scheme is presented to perform *in vitro-in vivo* extrapolations (ivive) (Section 7.3.1). This scheme is then used to derive  $k_m$  values using the experimental clearance values collected for human microsomes and hepatocytes. The extrapolated  $k_m$  values were compared to *in vivo* measurements in order to validate the ivive method in Section 7.3.2.

#### 7.3.1 *In vitro* to *in vivo* extrapolation scheme

The intrinsic hepatic clearance *in vitro* ( $CL_{INT,vitro}$ ) is calculated as the ratio between the  $V_{max}$  and  $K_m$  experimental values (valid when  $[S] < 10\%K_m$ ) [19].

The units of  $CL_{INT,vitro}$  depend on the *in vitro* system used:  $CL_{INT,vitro}$  is expressed as  $L \min^{-1} 10^{-6} \text{ cells}^{-1}$  for hepatocytes and as  $L \min^{-1} \text{ mg}_{PROT}^{-1}$  for liver microsomes or isolated enzymes. The procedure to perform ivive can be divided in 4 steps:

- 1)  $CL_{INT,vitro}$  is multiplied by the *in vitro* system scaling factor (SF) to obtain the intrinsic clearance in the liver ( $CL_{INT,liver}$ ,  $L \min^{-1} g_{LIV}^{-1}$ ):

$$CL_{INT,liver} = CL_{INT,vitro} \times SF \quad (\text{Eq. 7.1})$$

The *in vitro* system scaling factors are hepatocellularity for hepatocytes (HP,  $10^6 \text{ cells } g_{LIV}^{-1}$ ) and protein concentration for microsomes or isolated enzymes (PL,  $\text{mg}_{PROT} g_{LIV}^{-1}$ ).

- 2)  $CL_{INT,liver}$  is scaled to the intrinsic clearance for the whole-body ( $CL_{INT,vivo}$ ,  $L \min^{-1} kg_{BW}^{-1}$ ) via multiplication by liver weight (LW,  $g_{LIV} kg_{BW}^{-1}$ ):

$$CL_{INT,vivo} = CL_{INT,liver} \times LW \quad (\text{Eq. 7.2})$$

- 3) A physiological model of the liver is applied to obtain the total hepatic clearance ( $CL_H$ ,  $L \min^{-1} kg_{BW}^{-1}$ ). The most widely used model type is the 'well-stirred tank' model [19], which combines  $CL_{INT,vivo}$  with the hepatic blood flow ( $Q_H$ ,  $L \min^{-1} kg_{BW}^{-1}$ ) and a binding term ( $f_u$ , /) to obtain  $CL_H$ :

$$CL_H = \frac{Q_H \cdot f_u \cdot CL_{INT,vivo}}{Q_H + f_u \cdot CL_{INT,vivo}} \quad (\text{Eq. 7.3})$$

The parameter  $f_u$  is given by the ratio between unbound chemical fraction in blood plasma ( $f_{u,blood}$ ) and in the *in vitro* test system ( $f_{u,inc}$ ), assuming that only freely dissolved chemicals can be biotransformed. Studies showed that best predictions of *in vivo* clearances were obtained when disregarding  $f_u$  (i.e.  $f_u = 1$ ) for the extrapolation of *in vitro* hepatic clearances measured for diverse drugs in rat microsomes [157] and in human microsomes and hepatocytes [146, 158]. For this reason, Eq. 7.3 is rewritten as follows:

$$CL_H = \frac{Q_H \cdot CL_{INT,vivo}}{Q_H + CL_{INT,vivo}} \quad (\text{Eq. 7.3a})$$

- 4) Finally,  $CL_H$  is divided by the volume of distribution of the compound ( $V_d$ ,  $L kg_{BW}^{-1}$ ) and multiplied by a time conversion factor ( $1440 \text{ min d}^{-1}$ ) to calculate the biotransformation rate constants ( $k_m$ ,  $d^{-1}$ ):

$$k_m = \frac{CL_H}{V_d} \times 1440 \quad (\text{Eq. 7.4})$$

Table 7.2 lists the values of the biochemical and physiological parameters needed for the ivive (SF, LW,  $Q_H$ ). The SF values for the *in vitro* tests in humans were derived from meta-analysis [154]. When no meta-analysis data were available, SF parameters were calculated as arithmetic average of the values reported in the papers where  $K_m$  and  $V_{max}$  values were collected (Appendix G,



Table G1). The values of  $LW$  and  $Q_H$  are reference values taken from a study that gathered and averaged data from the scientific literature for various physiological parameters in mammals [155]. The  $V_d$  ( $L\ kg_{BW}^{-1}$ ) values can be estimated with the empirical equations in Table 7.3 [159], which depend on  $\log K_{ow}$  and charge state and were developed for drugs with human data.

Table 7.2. Biochemical and physiological parameters for ivive of liver clearance in humans, including the scaling factors for hepatocytes (HP), microsomes ( $PC_{micr}$ ) and the enzymes analysed in this thesis ( $PC_{CYP}$ ,  $PC_{ADH}$ ,  $PC_{ALDH}$  and  $PC_{FMO}$ ).

Description	Symbol	Units	Value	Source and comments
Hepatocellularity	HP	$10^6\ cells\ g_{LIV}^{-1}$	99	[154], meta-analysis
Protein content	$PC_{micr}$	$mg_{PROT}\ g_{LIV}^{-1}$	32	[154], meta-analysis
	$PC_{CYP}$	$mg_{PROT}\ g_{LIV}^{-1}$	32	[154], meta-analysis
	$PC_{ADH}$	$mg_{PROT}\ g_{LIV}^{-1}$	0.21	This thesis, average (Appendix G)
	$PC_{ALDH}$	$mg_{PROT}\ g_{LIV}^{-1}$	0.06	This thesis, average (Appendix G)
	$PC_{FMO}$	$mg_{PROT}\ g_{LIV}^{-1}$	0.13	This thesis, average pig and mouse (Appendix G)
Liver weight	LW	$g_{LIV}\ kg_{BW}^{-1}$	25.7	[155], average literature values
Hepatic blood flow	$Q_H$	$L\ min^{-1}\ kg_{BW}^{-1}$	0.021	[155], average literature values

Table 7.3.  $\log K_{ow}$ <sup>a</sup> dependent prediction of  $V_d$  ( $L\ kg_{BW}^{-1}$ ) at various predominant charge states at pH 7.4, taken from [159].

Predominant charge state at pH 7.4 <sup>b</sup>	$\log K_{ow}$ range <sup>c</sup>	Predicted $V_d$ ( $L\ kg_{BW}^{-1}$ )	Average fold error <sup>d</sup>
Uncharged (N)	$-3 < \log K_{ow} < 5$	1	2.8
Uncharged (N)	$5 \leq \log K_{ow} < 7$	10	4.3
Negatively charged (A)	$-2 < \log K_{ow} < 7$	0.2	2.5
Positively charged (B)	$-7 < \log K_{ow} \leq -2$	0.3	1.1
Positively charged (B)	$-2 < \log K_{ow} < 5$	$K_{ow}^{0.234} \cdot 10^{-0.0456}$	3.5
Positively charged (B)	$5 \leq \log K_{ow} < 8$	20	6.1

<sup>a</sup> $\log K_{ow}$  predicted with QSAR+ module of Cerius2 (v 4.6, Accelrys Inc, San Diego, USA).

<sup>b</sup>N = neutral, B = basic, A = acidic compounds. <sup>c</sup>From Table 10 and Figure 3 in [159].

<sup>d</sup>Fold Error = (exp.  $V_d$ /pred.  $V_d$ ) or (pred.  $V_d$ /exp.  $V_d$ ) whichever the greater.

### 7.3.2 Ivide for the data in this thesis for humans

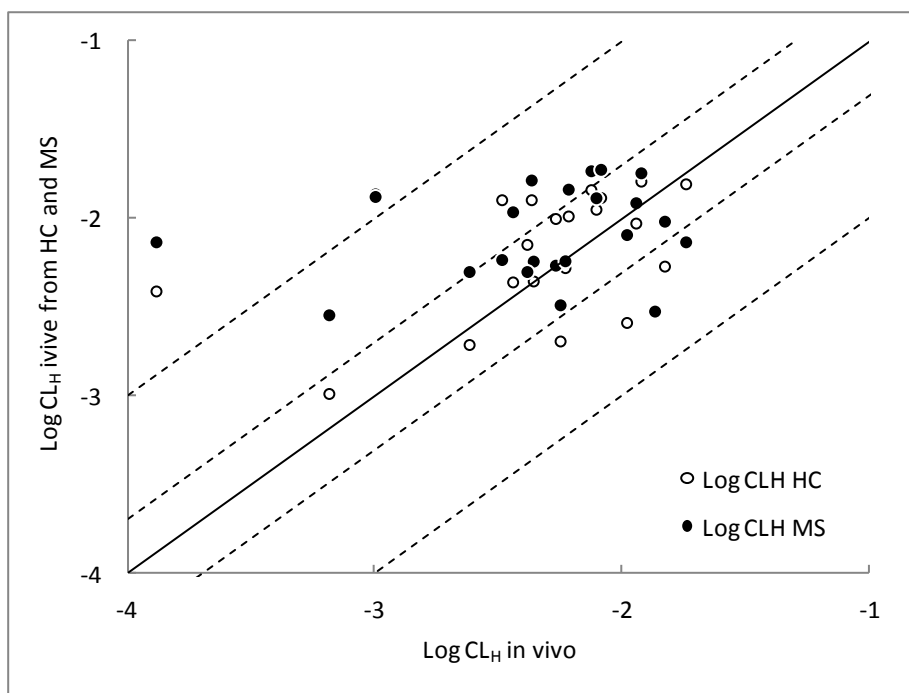
Here, the ivive method described in 7.3.1 is applied to a selection of compounds for which experimental  $CL_{INT}$  data were available for both microsomes and hepatocytes. The aim is to compare the  $CL_H$  values estimated with ivive to  $CL_H$  values measured *in vivo* for hepatic metabolism. For this reason, data collected for specific enzymes are not included in the ivive as they reflect only one pathway per substance, while the interest here is to obtain  $CL_H$  relative to the overall metabolism.

Measured *in vivo*  $CL_H$  values were retrieved from Paixão et al. 2010 [36], who calculated the clearance of 112 drugs from measured human intravenous *in vivo* pharmacokinetic data from Goodman et al. 2006 [160]. They collected data on intravenous total plasma clearance ( $CL_{total}$ ), fraction of drug eliminated by the kidneys, as well as oral bioavailability, and finally obtained hepatic *in vivo*  $CL_H$  values from  $CL_{total}$  by subtracting renal elimination routes and other non-hepatic ones.

Among the pharmaceuticals for which measured *in vivo*  $CL_H$  values were available, only the compounds included in both the hepatocytes (117 compounds) and microsomes (115 compounds) datasets were selected, for a total of 22 pharmaceuticals. The estimated *in vivo*  $CL_H$  of the 22 selected compounds were calculated by applying the ivive method in 7.3.1 to the experimental *in vitro*  $CL_{INT}$  collected for hepatocytes and microsomes for Chapter 6 (Appendix E).

The *in vitro*  $CL_H$  values estimated with the ivive are reported in Table G2 of Appendix G for human hepatocytes and microsomes, together with the measured *in vivo*  $CL_H$  values from Paixão et al. 2010 [36]. In Figure 7.1, the *in vivo*  $CL_H$  values estimated for hepatocytes and microsomes were plotted against the measured *in vivo*  $CL_H$  values.

Figure 7.1 (next page). Log transformed values of *in vivo*  $CL_H$  ( $L\ min^{-1}\ kg_{BW}^{-1}$ ) for humans measured from *in vivo* experiments plotted against the  $CL_H$  data calculated with the ivive described in Section 7.3.1 for hepatocytes (HC, white dots) and microsomes (MS, black dots). The black line represents the 1:1 line and the dotted lines the 10-fold and the 2-fold higher and lower intervals.



The Root Mean Squared Error (RMSE) and the percentage of predictions within 10-fold and 2-fold difference were used to evaluate the performance of the ivive method to predict *in vivo* CL<sub>H</sub> for hepatocytes and microsomes. The RMSE is calculated as the square root of the ratio between the sum of the square of all errors and the number of observations, and in this case it is expressed in Log units. The RMSE values of 0.51 for hepatocytes and 0.55 for microsomes indicate a low accuracy in predicting the *in vivo* clearances using ivive. When looking at the percentages of estimated values below 2-fold error (64% for hepatocytes and 50% for microsomes), the results of this synthesis were in accordance with previously described values for ivive with human hepatocytes with less than 50% of compounds within 2-fold error [36, 161, 162]. The reasons for lack of prediction may be either the low ability of CL<sub>INT</sub> data measured in *in vitro* assays to represent the *in vivo* situation or inappropriate ivive method. More investigations are needed to improve the accuracy of ivive methods.

In order to test whether the ivive has added value compared to the average of the *in vivo* measurements, the Coefficient of Efficiency (CoE) was calculated. The CoE is defined as one minus the ratio between the sum of the square of all errors and the variance of the observed values [163]:

$$\text{CoE} = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (\text{Eq. 7.6})$$

Where O is the observed value (i.e. the measured *in vivo* CL<sub>H</sub> value from Paixão et al. 2010 [36]) and P the predicted value (i.e. the *in vivo* CL<sub>H</sub> value estimated with ivive) for the 22 compounds. The negative CoE values obtained for hepatocytes (-0.1) and microsomes (-0.3) indicate that the average of the *in vivo* measurements over the chemicals is a better predictor compared to the ivive estimates for this set of chemicals. Two chemicals are mainly responsible for the negative CoE: gemfibrozil and atenolol. For these two compounds, *in vivo* CL<sub>H</sub> values estimated from hepatocytes and microsomes are more than one order of magnitude higher than the measured value. In addition, for this set of compounds the measured clearance rates are almost all within one or of magnitude, with the exception of atenolol. It is therefore recommended to repeat this analysis on a larger dataset with a wider range of clearance rates, when data will be available.

For the compounds analysed in this synthesis, clearances estimated using data from hepatocytes provide better results compared to microsomes data. This is probably because of the higher quality of the hepatocytes data, most of which were taken from controlled experiments. Moreover, isolated hepatocytes are liver cells, thus they contain the full complement of Phase 1 and Phase 2 metabolic enzymes and essential cofactors (e.g. NADPH). This means that all possible metabolic reactions can take place in hepatocytes and most transporter functions are preserved, better mimicking the *in vivo* systems [143]. Therefore, predictions of *in vivo* CL<sub>INT</sub> from hepatocytes data are usually more accurate than those from microsomal data [144], as microsomes assays provide exhaustive CL<sub>H</sub> values only when CYP metabolism is the dominant biotransformation pathway [143]. Most of CL<sub>H</sub> values were overestimated using ivive, i.e. 14 compounds for hepatocytes and 17 for microsomes. This is in contrast with previous studies, in which CL<sub>H</sub> in hepatocytes were generally underpredicted with ivive, for different reasons (e.g. neglect of extrahepatic metabolism, quality of cryopreserved hepatocytes, under-prediction potential of well-stirred model, etc.). The possible reasons for overestimation may be in the ivive method. The overestimation of CL<sub>H</sub>, and therefore the prediction of higher *k<sub>m</sub>* values, may lead to an underestimation of internal concentrations of chemicals. More studies are needed to determine the improvement of bioaccumulation models in mammals when biotransformation rates are included.

In conclusion, for the limited number of compounds analysed in this synthesis, the extrapolation of *in vitro* CL<sub>H</sub> from human hepatocytes provided *in vivo* CL<sub>H</sub> that were closer to the observed *in vivo* CL<sub>H</sub> compared to the microsomes results. This is in concordance with previous studies in which hepatocytes

generally provided more reliable estimation of the *in vivo* clearance due to their greater ability to mimic the *in vivo* situation. In order to have more extensive conclusions, additional data from *in vitro* experimental measurements are necessary for diverse compounds, for which *in vivo*  $CL_H$  values are already available in e.g. Paixão et al. 2010 [36].

#### 7.4 Change of hydrophobicity after metabolism

In Chapter 2, the change of  $K_{ow}$  after metabolism was quantified for parent compounds undergoing individual oxidation reactions catalysed by CYP, ADH and ALDH. For reactions metabolised by CYP, the  $K_{ow}$  of the metabolite was on average a factor of 10 lower if compared to the  $K_{ow}$  of its parent compound. For oxidations mediated by ALDH and ADH, the  $\text{Log } K_{ow}$  generally remained unchanged after metabolism. In a more recent and extensive study, Kirchmair et al. quantified the shift of  $\text{Log } K_{ow}$  for thousands of experimentally observed metabolic reactions of drugs as well as endobiotic compounds [164]. For drugs, the  $K_{ow}$  of the metabolites was on average a factor of 10 lower than the  $K_{ow}$  of the corresponding parent compound. This means that on average  $\text{Log } K_{ow,M} = \text{Log } K_{ow,P} - 1$ . This method provides information on the elimination rates of the metabolites in comparison to the parent compound. Metabolites are generally less hydrophobic (on average 10 times less), thus they are excreted faster as they would tend to accumulate to a lesser extent in fat.

A drawback of this estimation is that in reality it is very difficult to anticipate through which pathway(s) a compound is metabolised, sometimes even leading to metabolites with an increased  $K_{ow}$ . The main advantage of this method is that it allows to quantify the  $K_{ow}$  of metabolites based on the  $K_{ow}$  of the parent compound without knowing the identity (i.e. molecular structure) of the metabolites. This could be useful for risk assessment, as it is difficult to determine the molecular structure of metabolites for all parent compounds of interest. For example, the elimination constant of the metabolite ( $k_{ex}$ ,  $d^{-1}$ ) can be estimated from species weight and compound  $K_{ow}$  (equation by Hendriks et al. [15]) by using a  $K_{ow}$  of one order of magnitude lower than the  $K_{ow}$  of the parent compound. In combination with an estimated biotransformation rate for the parent compound (e.g. using *ivive* from a measured *in vitro*  $CL_{INT}$  or from a  $CL_{INT}$  predicted with the hepatocytes QSAR in Chapter 6), this would allow to estimate the accumulation of the parent compound and its metabolites, using a weight of evidence approach.

## 7.5 Conclusions

Substances that are taken up by organisms can be transformed through metabolic reactions, which contribute to their elimination. This process needs to be considered in the overall risk assessment, but the inclusion of metabolism in bioaccumulation models is still difficult. Biotransformation rates are difficult to obtain due to the complex processes involved, which depend on the distribution of the chemical and on the enzymatic action (binding to the enzyme and catalytic reaction). In addition, metabolic pathways are frequently not fully known and may differ depending on organisms and species. In order to understand better the processes influencing biotransformation, QSARs models were developed for metabolic constants in mammals, namely  $K_m$  and  $V_{max}$  of 4 oxidising enzymes (Chapters 3-5) and  $CL_{INT}$  of human hepatocytes and microsomes (Chapter 6).

The advantages and disadvantages of the models developed in Chapters 3 to 6 are also discussed in Chapter 7, with regard to the different descriptors and the different *in vitro* assays considered. While the QSARs for individual enzymes were helpful to interpret metabolic processes, their application to risk assessment is yet limited. Instead, the most promising results were obtained with human hepatocytes and microsomes. Especially for hepatocytes, the QSAR statistics are encouraging, allowing application of the outcomes *in vivo*. The performances of the QSARs are limited by the reliability of the *in vitro* assay systems [165]. The models can potentially be improved when more *in vitro* data become available from standardised experiments.

In addition, a general scheme for *in vitro* to *in vivo* extrapolation (*ivive*) was presented in Chapter 7 to estimate the biotransformation constant of chemicals needed for risk assessment. The *ivive* method was applied to derive  $k_m$  values using *in vitro* clearance values collected for human microsomes and hepatocytes in Chapter 6. The extrapolated  $k_m$  values were compared to *in vivo* measurement. The performances of the models were, however, limited by the reliability of the *in vitro* assay systems. The scheme needs to be validated on a wide array of chemicals, yet it could be useful for a first estimate of  $k_m$  in a weight of evidence approach.

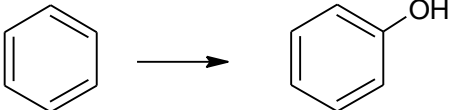
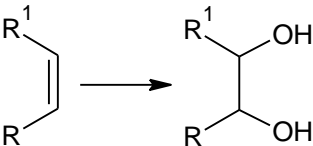
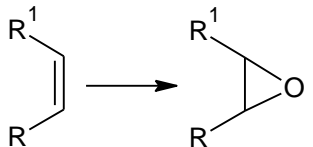
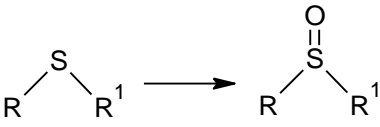

# **Appendix A**

## **Appendix to Chapter 2**

## Abbreviations

P450 = cytochrome P450 enzymes; ADH = alcohol dehydrogenase; ALDH = aldehyde dehydrogenase; PCBs = Polychlorinated biphenyls; PCDDs = Polychlorinated dibenzodioxins; PCDFs = Polychlorinated dibenzofurans; PBDEs = Polybrominated diphenyl ethers; PBBs = Polybromo biphenyls; PAHs = Polycyclic aromatic hydrocarbons; NHAs = Nitrogen heterocyclic aromatic compounds; OP = Organophosphorus; AA = Aromatic amines.

**Table A1.** Biotransformation reactions included in this study, with the typical classes of parent compounds and a representation of the reactions on chemical moieties.

Metabolic reaction	Chemical class of parent compounds	Reactions on chemical moieties
Reactions mediated by P450 enzymes		
Hydroxylation		
Aromatic	PCBs, aromatic hydrocarbons, heterocyclic compounds, PCDDs, PCDFs, PBDEs, PBBs.	
Aliphatic	Aliphatic hydrocarbons (alkanes and ketones), aromatic hydrocarbons, cyclic compounds, drugs (aliphatic amines, imides).	$R-CH_3 \longrightarrow R-CH_2OH$
Dihydroxylation	PAHs, NHAs, nitro PAHs, aromatic hydrocarbons, heterocyclic compounds.	
Epoxidation	PAH diols, NHA diols, aromatic hydrocarbons, heterocyclic compounds, aliphatic hydrocarbons (alkenes), cyclic alkenes, vinyl halides.	
Sulphoxidation	Thioethers (carbamate, thiocarbamate, OP pesticides).	
N-hydroxylation	AA (primary, secondary), heterocyclic AA (primary, secondary).	
$R^1=H$ for primary amines		



Reactions mediated by ADH enzymes		
Oxidation of primary alcohols to aldehydes	Aliphatic hydrocarbons (primary alcohols, allylic alcohols, glycols, glycol ethers, halohydrins), benzyl alcohols.	$\text{R}-\text{CH}_2\text{OH} \longrightarrow \text{R}-\text{CHO}$
Oxidation of secondary alcohols to ketones	Aliphatic hydrocarbons (secondary alcohols, allylic alcohols, cyclic compounds).	$\text{R}-\text{CH}(\text{OH})-\text{R}' \longrightarrow \text{R}-\text{C}(=\text{O})-\text{R}'$
Reaction mediated by ALDH enzymes		
Oxidation of aldehydes to acids	Aliphatic hydrocarbons (aldehydes), benzyl aldehydes.	$\text{R}-\text{CHO} \longrightarrow \text{R}-\text{COOH}$

**Table A2.** List of parent compounds (PC) and respective metabolites (M), divided by metabolic reaction. In the references, when the source is Toxin and Toxin Target Database (T3DB, <http://www.t3db.org/toxins/>), the T3DB ID is reported.

### 1. Hydroxylation mediated by P450

**1a. Aromatic hydroxylation by P450** (PCBs = Polychlorinated biphenyls; PCDDs = Polychlorinated dibenzodioxins; PCDFs = Polychlorinated dibenzofurans; PBDEs = Polybrominated diphenyl ethers; PBBs = Polybromo biphenyls)

PARENT COMPOUND (PC)				METABOLITE (M)			DIFFERENCE (ACD)	DIFFERENCE (exp)	REF.	
Class	Name	CAS number	Log $K_{ow}$ (ACD)	Log $K_{ow}$ (exp)	Name	CAS number	Log $K_{ow}$ (ACD)	Log $K_{ow}$ (exp)	$\frac{K_{ow}M}{\text{Log } K_{ow}P}$	Paper and/or T3DB ID
PCBs	4-chlorobiphenyl	2051-62-9	4.77	4.61	4'-chloro-4-biphenylol	28034-99-3	4.13	n.a.	-0.6	[166]
PCBs	3,3'-dichlorobiphenyl	2050-67-1	5.29	5.27	3,3'-dichloro-4-biphenylol	53890-78-1	4.66	n.a.	-0.6	[167]
PCBs	3,4-dichlorobiphenyl	2974-92-7	5.30	5.29	3,4-dichloro-4'-biphenylol	53890-77-0	4.66	n.a.	-0.6	[167]
PCBs	2,4,6-trichlorobiphenyl	35693-92-6	5.31	5.47	2,4,6-trichloro-4'-biphenylol	14962-28-8	4.68	n.a.	-0.6	[167]
PCBs	2,2',5,5'-tetrachlorobiphenyl	35693-99-3	5.83	6.09	2,2',5,5'-tetrachloro-4-biphenylol	51274-68-1	5.35	n.a.	-0.5	[168]
PCBs	3,3',4,4'-tetrachlorobiphenyl	32598-13-3	6.50	6.63	3,3',4',5'-tetrachloro-4-biphenylol	111810-41-4	5.88	n.a.	-0.6	[40, 169, 170]
PCBs	2,2',4,5,5'-pentachlorobiphenyl	37680-73-2	6.44	6.8	2,2',4,5,5'-pentachloro-4'-biphenylol	59512-50-4	5.96	n.a.	-0.5	[170]
PCBs	2,3,3',4,4'-pentachlorobiphenyl	11097-69-1	6.71	6.79	2',3,3',4',5'-pentachloro-4-biphenylol	149589-55-9	6.09	n.a.	-0.6	[171, 172]
PCBs	2,3',4,4',5-pentachlorobiphenyl	31508-00-6	6.77	7.12	2',3,3',4',5'-pentachloro-4-biphenylol	149589-55-9	6.09	n.a.	-0.7	[171, 172]

PCBs	3,3',4,4',5'-pentachlorobiphenyl	57465-28-8	7.03	n.a.	3,3',4,5,5'-pentachloro-4-biphenylol	130689-92-8	6.40	n.a.	-0.6	n.a.	[170, 173]
PCBs	2,2',3,4,4',5'-hexachlorobiphenyl	35065-28-2	6.98	7.44	2,2',3,4',5,5'-hexachloro-4-biphenylol	145413-90-7	6.50	n.a.	-0.5	n.a.	[171, 172]
PCBs	2,2',3,4',5,5'-hexachlorobiphenyl	51908-16-8	6.97	7.12	2,2',4,4',5,5'-hexachloro-3-biphenylol	54284-55-8	6.60	n.a.	-0.4	n.a.	[172]
PCBs	2,2',3,4',5,5'-hexachlorobiphenyl	51908-16-8	6.97	7.12	2,2',3,4',5,5'-hexachloro-4-biphenylol	145413-90-7	6.50	n.a.	-0.5	n.a.	[172]
PCBs	2,2',4,4',5,5'-hexachlorobiphenyl	35065-27-1	7.04	7.75	2,2',4,4',5,5'-hexachloro-3-biphenylol	54284-55-8	6.60	n.a.	-0.4	n.a.	[171, 172]
PCBs	2,2',4,4',5,5'-hexachlorobiphenyl	35065-27-1	7.04	7.75	2,2',3,4',5,5'-hexachloro-4-biphenylol	145413-90-7	6.50	n.a.	-0.5	n.a.	[171, 172]
PCBs	2,2',3,4',5,5',6'-heptachlorobiphenyl	52663-68-0	7.17	n.a.	2,2',3,4',5,5',6'-heptachloro-4-biphenylol	158076-68-7	6.85	n.a.	-0.3	n.a.	[171, 172]
PCBs	2,2',3,4,4',5',6'-heptachlorobiphenyl	52663-69-1	7.25	n.a.	2,2',3,4',5,5',6'-heptachloro-4-biphenylol	158076-68-7	6.85	n.a.	-0.4	n.a.	[172]
Aromatic hydrocarbons	benzene	71-43-2	2.18	2.13	phenol	108-95-2	1.54	1.46	-0.6	-0.7	T3D0006, [174]
Aromatic hydrocarbons	nitrobenzene	98-95-3	1.92	n.a.	4-nitrophenol	100-02-7	1.67	1.91	-0.3	n.a.	[175]
Aromatic hydrocarbons	2-ethylphenol	90-00-6	2.47	2.47	ethylhydroquinone	2349-70-4	1.51	n.a.	-1.0	n.a.	[176]
Aromatic hydrocarbons	4-ethylphenol	123-07-9	2.58	2.58	4-ethylcatechol	1124-39-6	1.84	n.a.	-0.7	n.a.	[176]
Aromatic hydrocarbons	4-(2,4,4-trimethyl-2-pentanyl)phenol	27193-28-8	5.18	n.a.	4-(2,4,4-trimethyl-2-pentanyl)-1,2-benzenediol	1139-46-4	4.39	n.a.	-0.8	n.a.	[177]

Continuation of 1a. Aromatic hydroxylation by P450

Aromatic hydrocarbons	phenol	108-95-2	1.54	1.46	hydroquinone	123-31-9	0.62	0.59	-0.9	T3D0182, [174]
Aromatic hydrocarbons	4-methylphenol	106-44-5	1.94	1.94	4-methyl-1,2-benzenediol	452-86-8	1.33	1.37	-0.6	[178]
Aromatic hydrocarbons	4-nitrophenol	100-02-7	1.67	1.91	4-nitro-1,2-benzenediol	3316-09-4	1.59	1.66	-0.3	[174, 179]
Aromatic hydrocarbons	2,4,5-trichlorophenol	95-95-4	3.84	3.72	3,4,6-trichloro-1,2-benzenediol	32139-72-3	3.44	3.60	-0.1	T3D0222
Aromatic hydrocarbons	aniline	62-53-3	1.14	0.90	4-aminophenol	123-30-8	0.01	0.04	-1.1	[174, 180]
Aromatic hydrocarbons	N-phenylacetamide	103-84-4	1.24	1.16	N-(4-hydroxyphenyl)acetamide	103-90-2	0.48	0.46	-0.7	[180]
Aromatic hydrocarbons	N-(3-chlorophenyl)acetamide	588-07-8	2.10	2.15	N-(3-chloro-4-hydroxyphenyl)acetamide	3964-54-3	1.35	0.91	-1.2	[180]
Aromatic hydrocarbons	N-(3-methylphenyl)acetamide	537-92-8	1.66	1.68	N-(4-hydroxy-3-methylphenyl)acetamide	16375-90-9	0.77	0.79	-0.9	[180]
Heterocyclic compounds	pyrazole	288-13-1	0.36	0.26	1H-pyrazol-4-ol	4843-98-5	-0.27	n.a.	-0.6	[174]
Heterocyclic compounds	3-pyridinol	109-00-2	0.05	0.48	5-hydroxy-2(1H)-pyridinone	5154-01-8	-0.40	n.a.	-0.4	[174]
Heterocyclic compounds	isoquinoline	119-65-3	2.02	2.08	4-isoquinolinol	3336-49-0	1.24	n.a.	-0.8	[181]
Heterocyclic compounds	benzo[f]quinoline	85-02-9	3.32	3.43	benzo[f]quinolin-7-ol	n.a.	2.68	n.a.	-0.6	[182]
Heterocyclic compounds	benzo[b]naphtho[2,1-d]thiophene	239-35-0	5.68	5.19	benzo[b]naphtho[2,1-d]thiophen-8-ol	n.a.	5.04	n.a.	-0.6	[183]
Heterocyclic compounds	7H-dibenzo[c,g]carbazole	194-59-2	6.12	6.40	7H-dibenzo[c,g]carbazol-5-ol	78448-06-3	5.49	n.a.	-0.6	[184]
Heterocyclic compounds	pyridine	110-86-1	0.84	0.65	4(1H)-pyridinone	108-96-3	0.22	-1.3	-0.6	[185]
PCDDs	2,8-dichlorodibenzo- <i>p</i> -dioxin	38964-22-6	5.59	n.a.	3-OH-2,8-dichlorodibenzo- <i>p</i> -dioxin	n.a.	5.27	n.a.	-0.3	[186]

PCDDs	2,3,7-trichlorodibenzo- <i>p</i> -dioxin	33857-28-2	5.94	n.a.	8-OH-2,3,7-trichlorodibenzo- <i>p</i> -dioxin	82019-04-3	5.63	n.a.	-0.3	n.a.	[187]
PCDDs	1,3,7,8-tetrachlorodibenzo- <i>p</i> -dioxin	50585-46-1	6.32	n.a.	2-OH-1,4,7,8-tetrachlorodibenzo- <i>p</i> -dioxin	n.a.	6.11	n.a.	-0.2	n.a.	[188]
PCDDs	2,3,7,8-tetrachlorodibenzo- <i>p</i> -dioxin	1746-01-6	6.29	6.80	2-OH-1,3,7,8-tetrachlorodibenzo- <i>p</i> -dioxin	82019-03-2	6.02	n.a.	-0.3	n.a.	[186]
PCDFs	2,3,7,8-tetrachlorodibenzo[b,d]fu ran	51207-31-9	6.51	n.a.	2,3,7,8-tetrachlorodibenzo[b,d]fu ran-4-ol	123566-86-9	6.03	n.a.	-0.5	n.a.	[186]
PBDEs	2,2',4,4'-tetrabromodiphenyl ether	5436-43-1	6.68	n.a.	3-OH-2,2',4,4'-tetrabromodiphenyl ether	n.a.	6.20	n.a.	-0.5	n.a.	[189]
PBDEs	2,2',4,4',5-pentabromodiphenyl ether	60348-60-9	7.31	n.a.	5'-OH-2,2',4,4',5-pentabromodiphenyl ether	n.a.	7.03	n.a.	-0.3	n.a.	[190]
PBBs	2-bromobiphenyl	2052-07-5	4.54	4.59	2-bromo-4-biphenylol	92-03-5	3.83	n.a.	-0.7	n.a.	[191]
PBBs	3-bromobiphenyl	2113-57-7	4.80	4.85	4'-bromo-4-biphenylol	29558-77-8	4.30	n.a.	-0.5	n.a.	[191]
PBBs	4,4'-dibromobiphenyl	92-86-4	5.79	5.72	4,4'-dibromo-3-biphenylol	n.a.	5.25	n.a.	-0.5	n.a.	[191]

### 1b. Aliphatic hydroxylation by P450

Aliphatic hydrocarbons	1,1,1-trichloroethane	71-55-6	2.35	2.49	2,2,2-trichloroethanol	115-20-8	0.97	1.42	-1.4	-1.1	[174]
Aliphatic hydrocarbons	hexane	110-54-3	3.76	3.9	2-hexanol	105-30-6	1.70	1.76	-2.1	-2.1	[192]
Aliphatic hydrocarbons	heptane	142-82-5	4.27	4.66	1-heptanol	111-70-6	2.37	2.62	-1.9	-2.0	[193]
Aliphatic hydrocarbons	2,6,10,14-tetramethylpentadecane	1921-70-6	9.76	n.a.	2,6,10,14-tetramethyl-2-pentadecanol	21980-66-5	7.93	n.a.	-1.8	n.a.	[194]
Aliphatic hydrocarbons	2-butanone	78-93-3	0.47	0.29	3-hydroxy-2-butanone	513-86-0	-0.30	n.a.	-0.8	n.a.	[195], p. 56
Aromatic hydrocarbons	toluene	108-88-3	2.72	2.73	phenylmethanol	100-51-6	1.06	1.10	-1.7	-1.6	[196]

## Continuation of 1b. Aliphatic hydroxylation by P450

Aromatic hydrocarbons	ethylbenzene	100-41-4	3.23	3.15	1-phenylethanol	1321-27-3	1.41	1.42	-1.8	-1.7	T3D009 9, [197]
Aromatic hydrocarbons	m-xylene	108-38-3	3.27	3.20	(3-methylphenyl)methanol	587-03-1	1.61	1.60	-1.7	-1.6	T3D005 8, [197]
Aromatic hydrocarbons	4-(2-nonyl)phenol	17404-66-9	6.04	n.a.	4-(8-hydroxynonan-2-yl)phenol	n.a.	3.98	n.a.	-2.1	n.a.	[198]
Aromatic hydrocarbons	4-nonylphenol	104-40-5	6.14	5.76	4-(9-hydroxynonyl)phenol	n.a.	4.24	n.a.	-1.9	n.a.	[199]
Cyclic compounds	camphor	76-22-2	2.09	2.38	3-hydroxy-camphor	10373-81-6	0.69	n.a.	-1.4	n.a.	[200];
Cyclic compounds	dodecylcyclohexane	1795-17-1	9.30	n.a.	12-cyclohexyl-2-dodecanol	n.a.	7.24	n.a.	-2.1	n.a.	[199]
Cyclic compounds	1,2,3,4-tetrahydronaphthalene	119-64-2	3.73	3.49	1,2,3,4-tetrahydro-1-naphthalenol	529-33-9	1.64	1.98	-2.1	-1.5	[201]
Aliphatic Amines	risperidone	106266-06-2	2.68	n.a.	9-hydroxy-risperidone	130049-84-2	1.41	n.a.	-1.3	n.a.	[202]
Aliphatic Amines	metoprolol	37350-58-6	1.63	1.88	1-hydroxy-metoprolol	110458-46-3	0.42	n.a.	-1.2	n.a.	[203]
Aliphatic Amines	perhexiline	6621-47-2	6.47	n.a.	cis-hydroxy-perhexiline	917877-73-7	4.91	n.a.	-1.6	n.a.	[203]
Aliphatic Amines	mexiletine	31828-71-4	2.12	2.15	6-hydroxymethylmexiletine	53566-98-6	0.57	n.a.	-1.6	n.a.	[204]
Imides	amobarbital	57-43-2	2.18	2.07	3'-hydroxyamobarbital	1421-07-4	0.32	n.a.	-1.9	n.a.	[205]

## 2. Epoxidation mediated by P450

## 2a. Aromatic epoxidation by P450 (PAHs = Polycyclic aromatic hydrocarbons; NHAs = Nitrogen heterocyclic aromatic)

PAH diols	7,8-dihydrobenzo[ <i>pqr</i> ]tetraphene-7,8-diol	13345-25-0	3.07	n.a.	7,8,9a-tetrahydrobenzo[1,12]tetrapheno[10,11-b]oxirene-7,8-diol	111137-80-5	1.70	n.a.	-1.4	n.a.	[206]
-----------	---	------------	------	------	---	-------------	------	------	------	------	-------

PAH diols	11,12-dihydronaphtho[1,2,3,4-pqr]tetraphene-11,12-diol	153857-27-3	4.25	n.a.	11,12,12a,13a-tetrahydronaphtho[4',3',2',1':1,12]tetrapheno[10,11-b]oxirene-11,12-diol	153857-28-4	2.88	n.a.	-1.4	n.a.	[206]
PAH diols	7,12-dimethyl-3,4-dihydro-3,4-tetraphenediolato	72617-60-8	3.54	n.a.	6,11-dimethyl-1a,2,3,11c-tetrahydrotetrapheno[1,2-b]oxirene-2,3-diol	74340-90-2	2.17	n.a.	-1.4	n.a.	[206]
NHA diols	3,4-dihydroadibenzo[a,j]acridin e-3,4-diol	117019-80-4	2.70	n.a.	1a,2,3,13c-tetrahydrobenzo[h][1]benzoxireno[2,3-a]acridine-2,3-diol	125276-72-4	1.46	n.a.	-1.2	n.a.	[43]
Aromatic hydrocarbons	bromobenzene	108-86-1	2.94	2.99	3-bromo-7-oxabicyclo[4.1.0]hept a-2,4-diene	51981-75-0	0.67	n.a.	-2.3	n.a.	[207]

## 2b. Aliphatic epoxidation by P450

Aromatic hydrocarbons	Styrene	100-42-5	2.82	2.95	2-phenyloxirane	20780-53-4	1.61	1.61	-1.2	-1.3	T3D0271, [208, 209]
Heterocyclic compounds	aflatoxin B1	1162-65-8	2.04	n.a.	aflatoxin B1 exo-8,9-epoxide	117859-29-7	1.62	n.a.	-0.4	n.a.	[210]
Heterocyclic compounds	2H-chromen-2-one	91-64-5	1.39	1.39	1a,7b-dihydro-2H-oxireno[c]chromen-2-one	143873-69-2	0.74	n.a.	-0.6	n.a.	[211]
Aliphatic hydrocarbons	Ethylene	74-85-1	1.32	1.13	oxirane	75-21-8	-0.58	-0.30	-1.9	-1.4	[210]
Aliphatic hydrocarbons	1-propene	115-07-1	1.83	1.77	2-methyloxirane	15448-47-2	0.33	0.03	-1.5	-1.7	[212]
Aliphatic hydrocarbons	2-butene	107-01-7	2.34	2.33	2,3-dimethyloxirane	1758-33-4	1.05	n.a.	-1.3	n.a.	[209]
Aliphatic hydrocarbons	1-octene	111-66-0	4.38	4.57	2-hexyloxirane	2984-50-1	2.88	n.a.	-1.5	n.a.	[213]

## Continuation of 2b. Aliphatic epoxidation by P450

Aliphatic hydrocarbons	Isoprene	78-79-5	2.35	2.42	2-methyl-2-vinyloxirane	1838-94-4	0.57	n.a.	-1.8	n.a.	[214]
Cyclic alkenes	Cyclohexene	110-83-8	2.92	2.86	7-oxabicyclo[4.1.0]heptane	137422-07-2	1.44	n.a.	-1.5	n.a.	[209]
Cyclic alkenes	bicyclo[2.2.1]hept-1-ene	21810-44-6	2.79	n.a.	2-oxatricyclo[3.2.1.0 <sup>1,3</sup> ]octane	n.a.	0.94	n.a.	-1.8	n.a.	[215]
Cyclic alkenes	4-vinylcyclohexene	100-40-3	3.72	3.93	3-vinyl-7-oxabicyclo[4.1.0]heptane	106-86-5	2.08	2.08	-1.6	-1.9	[216]
Vinyl halides	Chloroethene	75-01-4	1.69	n.a.	2-chlorooxirane	7763-77-1	-0.15	n.a.	-1.8	n.a.	[13, 210]
Vinyl halides	1,1-dichloroethene	75-35-4	1.77	2.13	2,2-dichlorooxirane	68226-83-5	0.44	n.a.	-1.3	n.a.	[210]
Vinyl halides	1,1,2-trichloroethene	79-01-6	2.57	2.42	2,2,3-trichlorooxirane	16967-79-6	1.17	n.a.	-1.4	n.a.	[210]
Vinyl halides	Aldrin	309-00-2	6.23	6.50	dieldrin	60-57-1	4.48	5.20	-1.8	-1.3	[217]
Aliphatic amines	N-methyl-N-nitrosoethenamine	4549-40-0	0.30	n.a.	N-methyl-N-nitrosooxiran-2-amine	n.a.	-0.82	n.a.	-1.1	n.a.	[210]
Aliphatic amides	Acrylamide	79-06-1	-0.56	-0.67	2-oxiranecarboxamide	5694-00-8	-1.08	n.a.	-0.5	n.a.	[218]
Vinyl nitriles	Acrylonitrile	107-13-1	0.17	0.25	2-oxiranecarbonitrile	4538-51-6	-0.98	n.a.	-1.2	n.a.	[218]
Esters	ethyl carbamate	51-79-6	-0.19	-0.15	2-oxiranyl carbamate	82617-23-0	-1.15	n.a.	-1.0	n.a.	[13, 210]
Esters	vinyl carbamate	15805-73-9	-0.20	n.a.	2-oxiranyl carbamate	82617-23-0	-1.15	n.a.	-0.9	n.a.	[219]

## 3. Dihydroxylation mediated by P450

## 3a. Aromatic dihydroxylation by P450 (PAHs = Polycyclic aromatic hydrocarbons; NHAs = Nitrogen heterocyclic aromatic)

NHAs	7-methylbenzo[c]acridine	3340-94-1	5.18	n.a.	7-methyl-8,9-dihydrobenzo[c]acridine-8,9-diol	n.a.	3.67	n.a.	-1.5	n.a.	[220]
------	--------------------------	-----------	------	------	---	------	------	------	------	------	-------



NHAs	dibenzo[a,j]acridine	224-42-0	5.82	n.a.	3,4-dihydrodibenz[a,j]acridine-3,4-diol	n.a.	4.49	n.a.	-1.3	n.a.	[220]
NHAs	quinoline	91-22-5	2.13	n.a.	5,6-dihydro-5,6-quinolinediol	87707-12-8	-0.43	n.a.	-2.6	n.a.	[181]
NHAs	benzo[h]quinoline	230-27-3	3.32	3.43	5,6-dihydrobenzo[h]quinoline-5,6-diol	87707-09-3	0.77	n.a.	-2.5	n.a.	[182]
PAHs	1H-indene	95-13-6	3.04	2.92	1,2-indanediol	46447-43-2	0.68	n.a.	-2.4	n.a.	[221]
PAHs	naphthalene	91-20-3	3.36	3.30	1,2-dihydro-1,2-naphthalenediol	7234-04-0	0.24	n.a.	-3.1	n.a.	[210]
PAHs	phenanthrene	85-01-8	4.55	4.46	9,10-dihydro-9,10-phenanthrenediol	25061-77-2	1.53	n.a.	-3.0	n.a.	[222]
PAHs	chrysene	218-01-9	5.73	5.81	1,2-dihydro-1,2-chrysenediol	28622-71-1	2.61	n.a.	-3.1	n.a.	[222]
PAHs	benzo[e]pyrene	192-97-2	6.19	6.44	4,5-dihydrobenzo[e]pyrene-4,5-diol	24961-49-7	3.00	n.a.	-3.2	n.a.	[223]
PAHs	pyrene	129-00-0	5.00	4.88	4,5-dihydro-4,5-pyrenediol	28622-70-0	1.81	n.a.	-3.2	n.a.	[224]
PAHs	tetraphene	56-55-3	5.73	5.76	3,4-dihydro-3,4-tetraphenediol	n.a.	4.40	n.a.	-1.3	n.a.	[225]
PAHs	benzo[k]tetraphene	53-70-3	6.91	6.75	benzo[k]tetraphene-1,2-diol	124027-77-6	5.58	n.a.	-1.3	n.a.	[226]
PAHs	anthracene	120-12-7	4.55	4.45	1,2-dihydro-1,2-anthracenediol	577-94-6	1.42	n.a.	-3.1	n.a.	[227]
Nitro PAHs	6-nitrochrysene	7496-02-8	5.47	n.a.	6-nitro-1,2-dihydro-1,2-chrysenediol	91828-72-7	2.35	n.a.	-3.1	n.a.	[228]
Nitro PAHs	1-nitrobenzo[ <i>pqr</i> ]tetraphene	70021-42-0	5.93	n.a.	1-nitro-7,8-dihydrobenzo[ <i>pqr</i> ]tetraphene-7,8-diol	88598-59-8	2.81	n.a.	-3.1	n.a.	[229]
Nitro PAHs	9-nitroanthracene	602-60-8	4.29	n.a.	9-nitro-1,2-dihydro-1,2-anthracenediol	n.a.	2.96	n.a.	-1.3	n.a.	[230]

Continuation of 3a. Aromatic dihydroxylation by P450 (PAHs = Polycyclic aromatic hydrocarbons; NHAs = Nitrogen heterocyclic aromatic)

Aromatic hydrocarbons	butylbenzene	104-51-8	4.28	4.38	2-butyl-1,3-benzenediol	13331-20-9	2.65	n.a.	-1.6	n.a.	[231]
Aliphatic Amines	propanolol	525-66-6	2.90	3.48	4,6-dihydroxypropanolol	114662-06-5	1.67	n.a.	-1.2	n.a.	[232]
Carbamates (Aromatic)	1-naphthyl methylcarbamate	63-25-2	2.34	2.36	5,6-dihydroxy-1-naphthyl methylcarbamate	24305-26-8	1.01	n.a.	-1.3	n.a.	[233]

### 3b. Aliphatic dihydroxylation by P450

Heterocyclic compounds	rotenone	83-79-4	4.03	n.a.	6',7'-dihydro-6',7'-dihydroxyrotenone	10585-57-6	2.41	n.a.	-1.6	n.a.	[234]
------------------------	----------	---------	------	------	---------------------------------------	------------	------	------	------	------	-------

### 4. Heteroatom oxygenation mediated by P450: Sulphoxidation (OP = Organophosphorus)

Carbamate pesticides	aldicarb	116-06-3	0.92	1.13	aldicarb sulphoxide	1646-87-3	-1.13	n.a.	-2.0	n.a.	[235]
Carbamate pesticides	methiocarb	2032-65-7	3.09	2.92	methiocarb sulphoxide	2635-10-1	0.34	n.a.	-2.7	n.a.	[236]
Carbamate pesticides	thiofanox	39196-18-4	2.39	2.75	thiofanox sulphoxide	39184-27-5	-0.25	n.a.	-2.6	n.a.	[237], p. 85
Carbamate pesticides	ethiofencarb	29973-13-5	2.04	2.04	ethiofencarb sulphoxide	336-34-5	-0.05	n.a.	-2.1	n.a.	[238]
Thiocarbamates	molinate	2212-67-1	2.67	n.a.	molinate sulphoxide	52236-29-0	0.70	n.a.	-2.0	n.a.	[239], p. 1348
Thiocarbamates	pebulate	1114-71-2	3.88	3.83	pebulate sulphoxide	51892-60-5	1.95	n.a.	-1.9	n.a.	[240], p. 261
Cyclohexanone pesticides	clethodim	99129-21-2	3.23	n.a.	clethodim sulphoxide	n.a.	1.00	n.a.	-2.2	n.a.	[241]
OP pesticides	fenthion	55-38-9	3.97	4.09	fenthion sulphoxide	3761-41-9	1.81	n.a.	-2.2	n.a.	[242]
OP pesticides	fenthion oxon	3254-63-5	2.44	n.a.	fenthion oxon sulphoxide	14086-35-2	-0.11	n.a.	-2.5	n.a.	[243]
OP pesticides	temephos	3383-96-8	5.96	5.96	temephos sulphoxide	17210-55-8	2.82	n.a.	-3.1	n.a.	[244], p. A-1

OP pesticides	demeton-S-methyl	919-86-8	1.60	1.02	oxydemeton methyl	301-12-2	-0.36	-0.74	-2.0	-1.8	[245], V7 p.808
OP pesticides	disulfoton	298-04-4	4.06	4.02	disulfoton sulphoxide	2497-07-6	1.79	1.73	-2.3	-2.3	[236]
OP pesticides	phorate	298-02-2	3.67	3.56	phorate sulphoxide	2588-03-6	1.82	1.78	-1.8	-1.8	[236]
OP pesticides	sulprofos	35400-43-2	4.55	5.48	sulprofos sulphoxide	n.a.	2.66	n.a.	-1.9	n.a.	[236]
OP pesticides	fenamiphos	22224-92-6	3.18	3.23	fenamiphos sulphoxide	31972-43-7	0.42	n.a.	-2.8	n.a.	[245], V7 p.848
OP pesticides	chlorthiophos	21923-23-9	5.38	n.a.	chlorthiophos sulphoxide	n.a.	3.51	n.a.	-1.9	n.a.	[246], p. 128
Triazine pesticides	ametryn	834-12-8	2.97	2.98	ametryn sulphoxide	80525-15-1	1.25	n.a.	-1.7	n.a.	[247]
Triazine pesticides	terbutryn	886-50-0	3.38	3.74	terbutryn sulphoxide	n.a.	1.66	n.a.	-1.7	n.a.	[247]
Heterocyclic compounds	albendazole	54965-21-8	2.91	n.a.	albendazole sulphoxide	54029-12-8	0.68	1.27	-2.2	n.a.	[13]

### 5. Heteroatom oxygenation mediated by P450: N-hydroxylation (AA = Aromatic amines)

AA, primary	mexiletine	31828-71-4	2.12	2.15	N-hydroxymexiletine	55304-17-1	2.85	n.a.	0.7	n.a.	[50]
AA, primary	dapsone	80-08-8	0.99	0.97	N-hydroxydapsone	32695-27-5	0.88	0.88	-0.1	-0.1	[50]
AA, primary	4-biphenylamine	92-67-1	2.89	2.86	N-hydroxy-4-biphenylamine	1204-79-1	2.43	n.a.	-0.5	n.a.	[248]
AA, primary	4,4'-biphenyldiamine	92-87-5	1.68	1.34	N-hydroxy-4,4'-biphenyldiamine	71609-27-3	1.23	n.a.	-0.5	n.a.	[249]
AA, primary	2-methyl-1-phenyl-2-propanamine	122-09-8	2.20	1.90	N-hydroxy-2-methyl-1-phenyl-2-propanamine	38473-30-2	2.46	n.a.	0.3	n.a.	[250]
AA, primary	4,4'-methylenedianiline	101-77-9	1.70	1.59	4-(4-aminobenzyl)-N-hydroxyaniline	n.a.	0.84	n.a.	-0.9	n.a.	[251]
AA, primary	4-[(E)-phenyldiazenyl]aniline	60-09-3	3.41	3.41	N-(4-phenylazophenyl)hydroxylamine	n.a.	2.98	n.a.	-0.4	n.a.	[252]
AA, secondary	N-ethylaniline	103-69-5	2.22	2.16	N-ethyl-N-hydroxyaniline	7447-59-8	1.72	n.a.	-0.5	n.a.	[253], p.198

Continuation of 5. Heteroatom oxygenation mediated by P450: N-hydroxylation (AA = Aromatic amines)

Polycyclic AA, primary	2-naphthalenamine	91-59-8	2.32	2.28	N-hydroxy-2-naphthalenamine	613-47-8	1.33	n.a.	-1.0	n.a.	[249]
Heterocyclic AA, primary	3-methyl-3H-imidazo[4,5-f]quinolin-2-amine	76180-96-6	1.41	1.46	N-hydroxy-3-methyl-3H-imidazo[4,5-f]quinolin-2-amine	77314-23-9	1.50	n.a.	0.1	n.a.	[254]
Heterocyclic AA, primary	3,5-dimethyl-3H-imidazo[4,5-f]quinolin-2-amine	77094-03-2	2.13	1.98	N-hydroxy-3,5-dimethyl-3H-imidazo[4,5-f]quinolin-2-amine	n.a.	2.27	n.a.	0.1	n.a.	[254]
Heterocyclic AA, primary	3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-2-amine	77500-04-0	1.13	1.01	N-hydroxy-3,8-dimethyl-3H-imidazo[4,5-f]quinoxalin-2-amine	115044-41-2	0.93	n.a.	-0.2	n.a.	[254]
Heterocyclic AA, primary	1-methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-amine	105650-23-5	2.44	2.23	N-hydroxy-1-methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-amine	124489-20-9	3.01	n.a.	0.6	n.a.	[254]
Heterocyclic AA, primary	1,4-dimethyl-5H-pyrido[4,3-b]indol-3-amine	62450-06-0	2.07	n.a.	N-hydroxy-1,4-dimethyl-5H-pyrido[4,3-b]indol-3-amine	n.a.	1.60	n.a.	-0.5	n.a.	[254]
Heterocyclic AA, primary	1-methyl-5H-pyrido[4,3-b]indol-3-amine	62450-07-1	2.26	n.a.	N-hydroxy-1-methyl-5H-pyrido[4,3-b]indol-3-amine	n.a.	1.95	n.a.	-0.3	n.a.	[254]
Heterocyclic AA, primary	6-methylpyrido[3',2':4,5]imidazo[1,2-a]pyridin-2-amine	67730-11-4	1.75	1.75	N-hydroxy-6-methylpyrido[3',2':4,5]imidazo[1,2-a]pyridin-2-amine	n.a.	1.02	n.a.	-0.7	n.a.	[254]
Heterocyclic AA, primary	9H-fluoren-2-amine	13924-50-0	3.15	n.a.	N-hydroxy-9H-fluoren-2-amine	53-94-1	2.30	n.a.	-0.9	n.a.	[254]
Heterocyclic AA, secondary	N-(9H-fluoren-2-yl)acetamide	53-96-3	3.26	n.a.	N-(9H-fluoren-2-yl)-N-hydroxyacetamide	53-95-2	3.07	n.a.	-0.2	n.a.	[50]

## 6. Oxidation mediated by ADH: Primary alcohol to aldehyde

Aliphatic hydrocarbons	methanol	67-56-1	-0.69	-0.77	formaldehyde	50-00-0	0.35	0.35	1.0	1.1	[255], p. 1886
Aliphatic hydrocarbons	ethanol	64-17-5	-0.18	-0.31	acetaldehyde	75-07-0	-0.11	-0.34	0.1	0.0	[255], p. 430
Aliphatic hydrocarbons	1-propanol	71-23-8	0.33	0.25	propionaldehyde	123-38-6	0.40	0.59	0.1	0.3	[256]
Aliphatic hydrocarbons	1-butanol	71-36-3	0.84	0.88	butyraldehyde	123-72-8	0.91	0.88	0.1	0.0	[195], p. 40
Aliphatic hydrocarbons	2-methyl-1-propanol	78-83-1	0.68	0.76	2-methylpropanal	78-84-2	0.76	n.a.	0.1	n.a.	[257], p.12
Aliphatic hydrocarbons	1-pentanol	71-41-0	1.35	1.51	valeraldehyde	110-62-3	1.42	n.a.	0.1	n.a.	[245], V6 p.429
Aliphatic hydrocarbons	1-hexanol	111-27-3	1.86	2.03	hexanal	66-25-1	1.93	1.78	0.1	-0.3	[245], V6 p.440
Aliphatic hydrocarbons	1-heptanol	111-70-6	2.37	2.62	heptanal	111-71-7	2.44	n.a.	0.1	n.a.	[245], V6 p.466
Aliphatic hydrocarbons	1-octanol	111-87-5	2.88	3.00	octanal	124-13-0	2.95	n.a.	0.1	n.a.	[258]
Aliphatic hydrocarbons	2-butoxyethanol	111-76-2	0.80	0.83	butoxyacetaldehyde	29043-89-8	0.51	n.a.	-0.3	n.a.	[259]
Aliphatic hydrocarbons	2-propen-1-ol	107-18-6	0.17	0.17	acrylaldehyde	107-02-8	0.26	-0.01	0.1	-0.2	[13]
Aliphatic hydrocarbons	(2E)-2-buten-1-ol	6117-91-5	0.69	n.a.	(2E)-2-butenal	4170-30-3	0.51	n.a.	-0.2	n.a.	[260]
Aliphatic hydrocarbons	3-methyl-2-buten-1-ol	556-82-1	1.06	n.a.	3-methyl-2-butenal	107-86-8	1.19	n.a.	0.1	n.a.	[261]
Aliphatic hydrocarbons	(2E,4E)-2,4-octadien-1-ol	18409-20-6	2.10	n.a.	(2E,4E)-2,4-octadienal	5577-44-6	2.40	n.a.	0.3	n.a.	[262]
Aliphatic hydrocarbons	(2E,4E)-2,4-decadien-1-ol	18409-21-7	3.36	n.a.	(2E,4E)-2,4-decadienal	2363-88-4	3.42	n.a.	0.1	n.a.	[262]
Aliphatic hydrocarbons	(E)-non-2-ene-1,4-diol	n.a.	1.60	n.a.	(2E)-4-hydroxy-2-nonenal	75899-68-2	1.90	n.a.	0.3	n.a.	[263]

## Continuation of 6. Oxidation mediated by ADH: Primary alcohol to aldehyde

Aliphatic hydrocarbons	ethylene glycol	107-21-1	-1.69	-1.36	glycolaldehyde	141-46-8	-1.60	n.a.	0.1	n.a.	[13]
Aliphatic hydrocarbons	propylene glycol	57-55-6	-1.34	-0.92	2-hydroxypropanal	3913-65-3	-1.25	n.a.	0.1	n.a.	[13]
Aliphatic hydrocarbons	2,2'-oxydiethanol	111-46-6	-1.51	n.a.	(2-hydroxyethoxy)acetaldehyde	17976-70-4	-1.67	n.a.	-0.2	n.a.	[264], p. 1232
Aliphatic hydrocarbons	2-methoxyethanol	109-86-4	-0.70	-0.77	methoxyacetaldehyde	10312-83-1	-1.02	n.a.	-0.3	n.a.	[265]
Aliphatic hydrocarbons	2-ethoxyethanol	110-80-5	-0.27	-0.32	2-ethoxyacetaldehyde	22056-82-2	-0.51	n.a.	-0.2	n.a.	[265]
Aliphatic hydrocarbons	2-butoxyethanol	111-76-2	0.80	0.83	butoxyacetaldehyde	29043-89-8	0.51	n.a.	-0.3	n.a.	[265]
Aliphatic hydrocarbons	2-chloroethanol	107-07-3	-0.08	0.03	chloroacetaldehyde	107-20-0	0.02	n.a.	0.1	n.a.	[266]
Aliphatic hydrocarbons	2-bromoethanol	1867-11-4	0.26	0.23	bromoacetaldehyde	17157-48-1	0.30	n.a.	0.0	n.a.	[267]
Aliphatic hydrocarbons	2-fluoroethanol	371-62-0	-0.40	-0.67	fluoroacetaldehyde	1544-46-3	-0.31	n.a.	0.1	n.a.	[268]
Aromatic Hydrocarbons	phenylmethanol	100-51-6	1.06	1.10	benzaldehyde	100-52-7	1.45	1.48	0.4	0.4	[245], V5 p.1039
Aromatic Hydrocarbons	(7-methoxy-1-naphthyl)methanol	n.a.	2.18	n.a.	7-methoxy-1-naphthaldehyde	n.a.	2.79	n.a.	0.6	n.a.	[269]
Aromatic Hydrocarbons	3-phenyl-1-propanol	122-97-4	1.88	1.88	3-phenylpropanal	104-53-0	1.78	n.a.	-0.1	n.a.	[270]
Aromatic Hydrocarbons	(2E)-3-phenyl-2-propen-1-ol	104-54-1	1.58	1.95	(2E)-3-phenylacrylaldehyde	104-55-2	2.12	1.90	0.5	-0.1	[271]
Aromatic Hydrocarbons	(6-methoxy-2-naphthyl)methanol	60201-22-1	2.18	n.a.	6-methoxy-2-naphthaldehyde	3453-33-6	2.63	n.a.	0.4	n.a.	[269]
Aromatic Hydrocarbons	4-(hydroxymethyl)-2-methoxyphenol	498-00-0	0.20	n.a.	4-hydroxy-3-methoxybenzaldehyde	121-33-5	1.21	1.21	1.0	n.a.	[270]

Aromatic Hydrocarbons	1-pyrenylmethanol	24463-15-8	3.88	n.a.	1-pyrenecarbaldehyde	3029-19-4	4.28	n.a.	0.4	n.a.	[54]
Aromatic Hydrocarbons	(8-methyl-1-pyrenyl)methanol	n.a.	4.45	n.a.	1-formyl-8-methylpyrene	n.a.	4.82	n.a.	0.4	n.a.	[54]

## 7. Oxidation mediated by ADH: Secondary alcohol to ketone

Aliphatic hydrocarbons	2-propanol	67-63-0	0.17	0.05	acetone	67-64-1	-0.04	-0.24	-0.2	-0.3	[195], p. 33
Aliphatic hydrocarbons	2-butanol	78-92-2	0.68	0.61	2-butanone	78-93-3	0.47	0.29	-0.2	-0.3	[195], p. 56
Aliphatic hydrocarbons	3-pentanol	584-02-1	1.19	1.21	3-pentanone	96-22-0	0.98	0.99	-0.2	-0.2	[272]
Aliphatic hydrocarbons	3-hexanol	623-37-0	1.70	1.65	3-hexanone	589-38-8	1.49	n.a.	-0.2	n.a.	[272]
Aliphatic hydrocarbons	4-heptanol	589-55-9	2.21	2.22	4-heptanone	123-19-3	2.00	n.a.	-0.2	n.a.	[272]
Aliphatic hydrocarbons	2-octanol	123-96-6	2.72	2.90	2-octanone	111-13-7	2.51	2.37	-0.2	-0.5	[273]
Aliphatic hydrocarbons	2-nonanol	628-99-9	3.23	n.a.	2-nonanone	821-55-6	3.02	3.14	-0.2	n.a.	[274]
Aliphatic hydrocarbons	3-methyl-2-butanol	598-75-4	1.04	1.28	3-methylbutan-2-one	563-80-4	0.82	0.84	-0.2	-0.4	[275]
Aliphatic hydrocarbons	3-butene-1,2-diol	497-06-3	-0.45	n.a.	1-hydroxy-3-buten-2-one	n.a.	-0.76	n.a.	-0.3	n.a.	[276]
Cyclic compounds	cyclohexanol	108-93-0	1.28	1.23	cyclohexanone	108-94-1	0.82	0.81	-0.5	-0.4	[270]

## 8. Oxidation mediated by ALDH: Aldehyde to acid

Aliphatic hydrocarbons	formaldehyde	50-00-0	0.35	0.35	formic acid	64-18-6	-0.54	-0.54	-0.9	-0.9	[255], p. 1886
Aliphatic hydrocarbons	acetaldehyde	1632-89-9	-0.11	-0.34	acetate	71-50-1	-0.32	-0.17	-0.2	0.2	[255], p. 430
Aliphatic hydrocarbons	propionaldehyde	123-38-6	0.40	0.59	propionic acid	79-09-4	0.19	0.33	-0.2	-0.3	[256]

Continuation of 8. Oxidation mediated by ALDH: Aldehyde to acid

Aliphatic hydrocarbons	butyraldehyde	123-72-8	0.91	0.88	butyrate	107-92-6	0.70	0.79	-0.2	-0.1	[195], p. 40
Aliphatic hydrocarbons	2-methylpropanal	26140-46-5	0.76	n.a.	2-methylpropanoic acid	79-31-2	0.54	0.94	-0.2	n.a.	[257], p.12
Aliphatic hydrocarbons	valeraldehyde	110-62-3	1.42	n.a.	valeric acid	109-52-4	1.21	1.39	-0.2	n.a.	[245], V6 p.429
Aliphatic hydrocarbons	hexanal	66-25-1	1.93	1.78	hexanoic acid	142-62-1	1.72	1.92	-0.2	0.1	[245], V6 p.440
Aliphatic hydrocarbons	heptanal	111-71-7	2.44	n.a.	heptanoic acid	111-14-8	2.23	2.42	-0.2	n.a.	[245], V6 p.466
Aliphatic hydrocarbons	octanal	124-13-0	2.95	n.a.	octanoic acid	124-07-2	2.74	3.05	-0.2	n.a.	[258]
Aliphatic hydrocarbons	butoxyacetaldehyde	29043-89-8	0.51	n.a.	butoxyacetic acid	2516-93-0	0.59	n.a.	0.1	n.a.	[259]
Aliphatic hydrocarbons	acrylaldehyde	107-02-8	0.26	-0.01	acrylic acid	79-10-7	0.15	0.35	-0.1	0.4	[277]
Aliphatic hydrocarbons	(2E)-2-butenal	4170-30-3	0.51	n.a.	(2E)-2-butenic acid	107-93-7	0.66	0.72	0.1	n.a.	[278]
Aliphatic hydrocarbons	(2E,4E)-2,4-octadienal	30361-28-5	2.40	n.a.	(2E,4E)-2,4-octadienoic acid	83615-26-3	2.29	n.a.	-0.1	n.a.	[262]
Aliphatic hydrocarbons	(2E,4E)-2,4-decadienal	2363-88-4	3.42	n.a.	(2E,4E)-2,4-decadienoic acid	30361-33-2	3.31	n.a.	-0.1	n.a.	[262]
Aliphatic hydrocarbons	(2E)-4-hydroxy-2-nonenal	75899-68-2	1.90	n.a.	(2E)-4-hydroxy-2-nonenic acid	n.a.	1.88	n.a.	0.0	n.a.	[263]
Aliphatic hydrocarbons	glycolaldehyde	141-46-8	-1.60	n.a.	glycol acid	79-14-1	-1.20	-1.11	0.4	n.a.	[13]
Aliphatic hydrocarbons	2-hydroxypropanal	3913-65-3	-1.25	n.a.	2-hydroxypropanoate	113-21-3	-0.85	-0.72	0.4	n.a.	[13]
Aliphatic hydrocarbons	(2-hydroxyethoxy)acetaldehyde	17976-70-4	-1.67	n.a.	(2-hydroxyethoxy)acetate	n.a.	-1.51	n.a.	0.2	n.a.	[264], p. 1232
Aliphatic hydrocarbons	methoxyacetaldehyde	10312-83-1	-1.02	n.a.	methoxyacetic acid	625-45-6	-0.94	n.a.	0.1	n.a.	[265]



Aliphatic hydrocarbons	2-ethoxyacetaldehyde	22056-82-2	-0.51	n.a.	ethoxyacetic acid	627-03-2	-0.43	n.a.	0.1	n.a.	[265]
Aliphatic hydrocarbons	butoxyacetaldehyde	29043-89-8	0.51	n.a.	butoxyacetic acid	2516-93-0	0.59	n.a.	0.1	n.a.	[265]
Aliphatic hydrocarbons	chloroacetaldehyde	107-20-0	0.02	n.a.	chloroacetic acid	79-11-8	-0.05	0.22	-0.1	n.a.	[266]
Aliphatic hydrocarbons	bromoacetaldehyde	17157-48-1	0.30	n.a.	bromoacetic acid	79-08-3	0.51	0.41	0.2	n.a.	[267]
Aliphatic hydrocarbons	fluoroacetaldehyde	1544-46-3	-0.31	n.a.	fluoroacetic acid	144-49-0	-0.23	n.a.	0.1	n.a.	[268]
Aromatic hydrocarbons	benzaldehyde	100-52-7	1.45	1.48	benzoic acid	65-85-0	1.56	1.87	0.1	0.4	[70]
Aromatic hydrocarbons	7-methoxy-1-naphthaldehyde	n.a.	2.79	n.a.	7-methoxy-1-naphthoic acid	7498-58-0	2.74	n.a.	0.0	n.a.	[269]
Aromatic hydrocarbons	3-phenylpropanal	104-53-0	1.78	n.a.	3-phenylpropanoic acid	501-52-0	1.84	1.84	0.1	n.a.	[70]
Aromatic hydrocarbons	(2E)-3-phenylacrylaldehyde	104-55-2	2.12	1.90	(2E)-3-phenylacrylic acid	621-82-9	2.41	2.13	0.3	0.2	[70]
Aromatic hydrocarbons	6-methoxy-2-naphthaldehyde	3453-33-6	2.63	n.a.	6-methoxy-2-naphthoic acid	2471-70-7	2.74	n.a.	0.1	n.a.	[269]
Aromatic hydrocarbons	4-hydroxy-3-methoxybenzaldehyde	121-33-5	1.21	1.21	4-hydroxy-3-methoxybenzoic acid	121-34-6	1.30	1.43	0.1	0.2	[70]
Aromatic hydrocarbons	1-pyrenecarbaldehyde	3029-19-4	4.28	n.a.	1-pyrenecarboxylic acid	19694-02-1	4.39	n.a.	0.1	n.a.	[279]
Aromatic hydrocarbons	1-formyl-8-methylpyrene	n.a.	4.82	n.a.	8-methyl-1-pyrene-1-carboxylic acid	n.a.	4.93	n.a.	0.1	n.a.	[279]



## **Appendix B**

### **Appendix with data sets for Chapters 3-5**

**Formulas of the statistical coefficients used as measures of model fitting and predictive power.**

**Coefficient of determination ( $R^2$ )**

$$R^2 = 1 - (RSS/SS_y)$$

The Residual Sum of Squares (RSS) is the sum of the squared difference between the experimental response ( $y$ ) and the response calculated by the model ( $\hat{y}$ ):

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

The total Sum of Squares ( $SS_y$ ) is the sum of the squared differences between the experimental response ( $y$ ) and the average experimental response ( $\bar{y}$ ):

$$SS_y = \sum_{i=1}^n (y_i - \bar{y})^2$$

**Adjusted coefficient of determination ( $R^2_{adj}$ )**

$$R^2_{adj} = 1 - (1 - R^2)(n - 1)/(n - p)$$

where  $n$  is the number of compounds in the dataset and  $p$  is the number of variables.

**Root Mean Squared Error (RMSE)**

$$RMSE = \sqrt{RSS/n}$$

where  $n$  is the number of compounds in the dataset.

**Leave-one-out cross-validated  $R^2$  ( $Q^2_{loo}$ )**

$$Q^2_{loo} = 1 - (PRESS/SS_y)$$

The Predictive Error Sum of Squares (PRESS) is the sum of the squared differences between the experimental response ( $y$ ) and the response predicted by the model for the object that was not used for model estimation ( $y_{i/i}$ ):

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_{i/i})^2$$

where the notation  $i/i$  indicates that the response is predicted by a model estimated when the  $i$ -th compound was left out from the training set.

**RMSE of the Leave-one-out cross-validation (RMSE<sub>LOO</sub>)**

$$\text{RMSE}_{\text{LOO}} = \sqrt{\text{PRESS}/n}$$

where n is the number of compounds in the dataset.

**Akaike's information criterion (AIC)**

$$\text{AIC} = \text{RSS} + (n + p')/(n - p')^2$$

where p' is the number of variables plus one and n is the number of compounds.

**Table B1.** Original data (na = CAS number value not available).

ADH

Ref.	Species	Isoenz	pH	T	Compound name	CAS	K <sub>m</sub> , $\mu\text{M}$	V <sub>max</sub>	V <sub>max</sub> units	k <sub>cat</sub> , $\text{min}^{-1}$	V <sub>max</sub> , $\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$
[280]	Human	ADH2	7.5	25	ethanol	64-17-5	34000	0.50	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.50
[280]	Human	ADH2	7.5	25	acetaldehyde	75-07-0	30000	21	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		21.00
[281]	Human	ADH1	7.5	25	ethanol	64-17-5	4200	27	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$		0.68
[281]	Human	ADH1	7.5	25	ethanol	64-17-5	49	9.2	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$		0.23
[281]	Human	ADH1	10	25	ethanol	64-17-5	1500	150	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$		3.75
[281]	Human	ADH1	10	25	ethanol	64-17-5	1600	18	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$		0.45
[281]	Human	ADH1	10	25	ethanol	64-17-5	3200	220	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$		5.50
[281]	Human	ADH1	10	25	ethanol	64-17-5	1700	120	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$		3.00
[282]	Human	ADH1	7.5	25	ethanol	64-17-5	35909	7.90	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		7.90
[282]	Human	ADH1	7.5	25	propanol	71-23-8	17241	5.00	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		5.00
[282]	Human	ADH1	7.5	25	butanol	71-36-3	4082	4.00	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		4.00
[282]	Human	ADH1	7.5	25	pentanol	71-41-0	2556	4.60	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		4.60
[282]	Human	ADH1	7.5	25	hexanol	111-27-3	1000	4.60	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		4.60
[282]	Human	ADH1	7.5	25	2-butanol	78-92-2	66667	0.40	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.40
[282]	Human	ADH1	7.5	25	isobutanol	78-83-1	44262	2.70	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		2.70
[282]	Human	ADH1	7.5	25	isopentanol	123-51-3	2615	3.40	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		3.40
[282]	Human	ADH1	7.5	25	ethanol	64-17-5	860	8.60	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		8.60
[282]	Human	ADH1	7.5	25	propanol	71-23-8	592	7.10	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		7.10
[282]	Human	ADH1	7.5	25	butanol	71-36-3	338	9.80	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		9.80
[282]	Human	ADH1	7.5	25	pentanol	71-41-0	131	6.80	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		6.80

[282]	Human	ADH1	7.5	25	hexanol	111-27-3	40	6.80	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		6.80
[282]	Human	ADH1	7.5	25	2-butanol	78-92-2	17895	3.40	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		3.40
[282]	Human	ADH1	7.5	25	isobutanol	78-83-1	3667	6.60	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		6.60
[282]	Human	ADH1	7.5	25	isopentanol	123-51-3	200	7.80	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		7.80
[282]	Human	ADH1	7.5	25	methanol	67-56-1	3000000	3.60	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		3.60
[282]	Human	ADH1	7.5	25	ethanol	64-17-5	22	0.09	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.09
[282]	Human	ADH1	7.5	25	propanol	71-23-8	19	0.07	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.07
[282]	Human	ADH1	7.5	25	butanol	71-36-3	12	0.07	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.07
[282]	Human	ADH1	7.5	25	pentanol	71-41-0	19	0.09	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.09
[282]	Human	ADH1	7.5	25	hexanol	111-27-3	22	0.11	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.11
[282]	Human	ADH1	7.5	25	2-butanol	78-92-2	333	0.08	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.08
[282]	Human	ADH1	7.5	25	isobutanol	78-83-1	61	0.10	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.10
[282]	Human	ADH1	7.5	25	isopentanol	123-51-3	14	0.07	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.07
[282]	Human	ADH1	7.5	25	methanol	67-56-1	11020	0.05	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.05
[282]	Human	ADH1	7.5	25	cyclohexanol	108-93-0	3895	0.07	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.07
[68]	Human	ADH1	7	25	acetaldehyde	75-07-0	2000			350	4.38
[68]	Human	ADH1	7	25	pentanal	110-62-3	8			450	5.63
[68]	Human	ADH1	7	25	benzaldehyde	100-52-7	45			670	8.38
[68]	Human	ADH1	7	25	cyclohexanone	108-94-1	230			590	7.38
[68]	Human	ADH1	7	25	acetaldehyde	75-07-0	76			380	4.75
[68]	Human	ADH1	7	25	pentanal	110-62-3	16			280	3.50
[68]	Human	ADH1	7	25	octanal	124-13-0	1			300	3.75
[68]	Human	ADH1	7	25	benzaldehyde	100-52-7	380			190	2.38
[68]	Human	ADH1	7	25	cyclohexanone	108-94-1	110			21	0.26
[68]	Human	ADH2	7	25	acetaldehyde	75-07-0	8300			650	8.13

[68]	Human	ADH2	7	25	pentanal	110-62-3	53	920	11.50
[68]	Human	ADH2	7	25	octanal	124-13-0	6.8	1200	15.00
[68]	Human	ADH2	7	25	benzaldehyde	100-52-7	10	450	5.63
[68]	Human	ADH2	7	25	cyclohexanone	108-94-1	100	20	0.25
[68]	Human	ADH3	7	25	octanal	124-13-0	75	75	0.94
[270]	Human	ADH2	10	25	ethanol	64-17-5	120000	470	5.88
[270]	Human	ADH2	10	25	pentanol	71-41-0	90	480	6.00
[270]	Human	ADH2	10	25	octanol	111-87-5	7	500	6.25
[270]	Human	ADH2	10	25	benzyl alcohol	100-51-6	7	550	6.88
[270]	Human	ADH2	10	25	3-phenyl-1-propanol	122-97-4	32	450	5.63
[270]	Human	ADH2	10	25	vanillyl alcohol	498-00-0	34	520	6.50
[270]	Human	ADH2	10	25	tryptophol	526-55-6	200	110	1.38
[270]	Human	ADH2	10	25	12-hydroxydodecanoic acid	505-95-3	230	240	3.00
[270]	Human	ADH2	10	25	16-hydroxyhexadecanoic acid	506-13-8	60	370	4.63
[270]	Human	ADH2	10	25	ethylene glycol	107-21-1	290000	45	0.56
[270]	Human	ADH2	10	25	2-propanol	67-63-0	560000	45	0.56
[270]	Human	ADH2	10	25	cyclohexanol	108-93-0	210000	35	0.44
[270]	Human	ADH2	10	25	2-deoxy-d-ribose	533-67-5	310000	170	2.13
[283]	Horse	ADH1	10	25	ethanol	64-17-5	2200	500	6.25
[283]	Horse	ADH1	10	25	propanol	71-23-8	980	520	6.50
[283]	Horse	ADH1	10	25	butanol	71-36-3	290	290	3.63
[283]	Horse	ADH1	10	25	pentanol	71-41-0	240	380	4.75
[283]	Horse	ADH1	10	25	hexanol	111-27-3	130	330	4.13
[283]	Horse	ADH1	10	25	octanol	111-87-5	27	260	3.25
[283]	Horse	ADH1	10	25	cyclohexanol	108-93-0	6100	820	10.25



[283]	Horse	ADH1	10	25	12-hydroxydodecanoic acid	505-95-3	270	110	1.38
[283]	Horse	ADH1	10	25	16-hydroxyhexadecanoic acid	506-13-8	18	100	1.25
[283]	Human	ADH1	10	25	ethanol	64-17-5	1200	35	0.44
[283]	Human	ADH1	10	25	propanol	71-23-8	720	38	0.48
[283]	Human	ADH1	10	25	butanol	71-36-3	210	33	0.41
[283]	Human	ADH1	10	25	pentanol	71-41-0	150	38	0.48
[283]	Human	ADH1	10	25	hexanol	111-27-3	57	36	0.45
[283]	Human	ADH1	10	25	octanol	111-87-5	13	25	0.31
[283]	Human	ADH1	10	25	cyclohexanol	108-93-0	14500	14	0.18
[283]	Human	ADH1	10	25	12-hydroxydodecanoic acid	505-95-3	85	28	0.35
[283]	Human	ADH1	10	25	16-hydroxyhexadecanoic acid	506-13-8	24	18	0.23
[283]	Human	ADH1	10	25	ethanol	64-17-5	1100	140	1.75
[283]	Human	ADH1	10	25	propanol	71-23-8	440	290	3.63
[283]	Human	ADH1	10	25	butanol	71-36-3	160	350	4.38
[283]	Human	ADH1	10	25	pentanol	71-41-0	160	370	4.63
[283]	Human	ADH1	10	25	hexanol	111-27-3	50	310	3.88
[283]	Human	ADH1	10	25	octanol	111-87-5	10	160	2.00
[283]	Human	ADH1	10	25	cyclohexanol	108-93-0	42	130	1.63
[283]	Human	ADH1	10	25	12-hydroxydodecanoic acid	505-95-3	66	160	2.00
[283]	Human	ADH3	10	25	butanol	71-36-3	1000	150	1.88
[283]	Human	ADH3	10	25	pentanol	71-41-0	22000	240	3.00
[283]	Human	ADH3	10	25	hexanol	111-27-3	8200	130	1.63
[283]	Human	ADH3	10	25	octanol	111-87-5	1200	220	2.75

[283]	Human	ADH3	10	25	12-hydroxydodecanoic acid	505-95-3	60			170	2.13
[260] <sup>2</sup>	Horse	ADH1	7.3	37	ethanol	64-17-5	2460	0.013 0	$\mu\text{mol min}^{-1} \text{mg}^{-1}$		0.01
[260]	Horse	ADH1	7.3	37	allyl alcohol	107-18-6	390	0.003 2	$\mu\text{mol min}^{-1} \text{mg}^{-1}$		0.00
[260]	Horse	ADH1	7.3	37	2-buten-1-ol	6117-91-5	710	0.002 3	$\mu\text{mol min}^{-1} \text{mg}^{-1}$		0.00
[284]	Human	ADH1	10.5	25	ethanol	64-17-5	1800	0.07	$\mu\text{mol min}^{-1} \text{mg}^{-1}$		0.07
[284]	Rat	ADH1	10.5	25	ethanol	64-17-5	800	0.02	$\mu\text{mol min}^{-1} \text{mg}^{-1}$		0.02
[285]	Rat	ADH3	10	25	pentanol	71-41-0	78000			212	2.65
[285]	Rat	ADH3	10	25	octanol	111-87-5	510			140	1.75
[285]	Rat	ADH3	10	25	12-hydroxydodecanoic acid	505-95-3	100			230	2.88
[285]	Rat	ADH3	10	25	2-buten-1-ol	6117-91-5	60000			350	4.38
[285]	Rat	ADH3	10	25	cyclohexanol	108-93-0	1900000			23	0.29
[285]	Rat	ADH1	10	25	methanol	67-56-1	380000			12	0.15
[285]	Rat	ADH1	10	25	ethanol	64-17-5	1400			60	0.75
[285]	Rat	ADH1	10	25	butanol	71-36-3	170			50	0.63
[285]	Rat	ADH1	10	25	pentanol	71-41-0	80			70	0.88
[285]	Rat	ADH1	10	25	octanol	111-87-5	25			60	0.75
[285]	Rat	ADH1	10	25	12-hydroxydodecanoic acid	505-95-3	13			76	0.95
[285]	Rat	ADH1	10	25	2-buten-1-ol	6117-91-5	350			100	1.25
[285]	Rat	ADH1	10	25	cyclohexanol	108-93-0	2200			90	1.13

<sup>2</sup> The  $V_{\max}$  data from ref. 260.Fontaine F.R., et al., 2002. *Oxidative bioactivation of crotyl alcohol to the toxic endogenous aldehyde crotonaldehyde: Association of protein carbonylation with toxicity in mouse hepatocytes*. Chemical Research in Toxicology, 15(8): p. 1051-1058.  
 And from ref. 284. Herrera E., et al., 1983. *Comparative kinetics of human and rat liver alcohol dehydrogenase*. Biochemical Society Transactions, 11: p. 729-730. were not inserted in the regressions, see Table B2.

[285]	Rat	ADH1	10	25	benzyl alcohol	100-51-6	50		60	0.75
[285]	Rat	ADH3	7.5	25	octanol	111-87-5	1600		4	0.05
[285]	Rat	ADH3	7.5	25	octanal	124-13-0	3500		250	3.13
[285]	Rat	ADH1	7.5	25	ethanol	64-17-5	1400		39	0.49
[285]	Rat	ADH1	7.5	25	octanol	111-87-5	60		13	0.16
[285]	Rat	ADH1	7.5	25	acetaldehyde	75-07-0	170		170	2.13
[285]	Rat	ADH1	7.5	25	butanal	123-72-8	90		440	5.50
[285]	Rat	ADH1	7.5	25	octanal	124-13-0	50		440	5.50
[285]	Rat	ADH1	7.5	25	m-nitrobenzaldehyde	99-61-6	60		260	3.25
[286]	Human	ADH1	10	25	methanol	67-56-1	2900		20	0.25
[286]	Human	ADH1	10	25	ethanol	64-17-5	2700		240	3.00
[286]	Human	ADH1	10	25	propanol	71-23-8	400		250	3.13
[286]	Human	ADH1	10	25	butanol	71-36-3	33		240	3.00
[286]	Human	ADH1	10	25	pentanol	71-41-0	40		240	3.00
[286]	Human	ADH1	10	25	hexanol	111-27-3	86		250	3.13
[286]	Human	ADH1	6.8	25	formaldehyde	50-00-0	24000		470	5.88
[286]	Human	ADH1	6.8	25	acetaldehyde	75-07-0	1400		6940	86.75
[286]	Human	ADH1	6.8	25	propanal	123-38-6	100		6940	86.75
[286]	Human	ADH1	6.8	25	butanal	123-72-8	22		6180	77.25
[286]	Human	ADH1	6.8	25	pentanal	110-62-3	31		6940	86.75
[286]	Human	ADH1	6.8	25	hexanal	66-25-1	64		6940	86.75
[287]	Horse	ADH1	7	25	ethanol	64-17-5	40000	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	76.2	1.91
[287]	Horse	ADH1	7	25	propanol	71-23-8	2900	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	69.6	1.74
[287]	Horse	ADH1	7	25	butanol	71-36-3	91	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	56.4	1.41
[287]	Horse	ADH1	7	25	hexanol	111-27-3	30	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	59.4	1.49

[287]	Horse	ADH1	7	25	benzyl alcohol	100-51-6	46	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	46.8	1.17
[287]	Horse	ADH1	7	25	acetaldehyde	75-07-0	6000	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	1908	47.70
[287]	Horse	ADH1	7	25	propanal	123-38-6	1240	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	2514	62.85
[287]	Horse	ADH1	7	25	butanal	123-72-8	60	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	2124	53.10
[287]	Horse	ADH1	7	25	hexanal	66-25-1	12	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	3396	84.90
[287]	Horse	ADH1	7	25	benzaldehyde	100-52-7	148	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	1770	44.25
[287]	Horse	ADH1	7	25	3-oxo-5 $\beta$ -androstane-17 $\beta$ -ol	1239-31-2	31	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	123	3.08
[287]	Horse	ADH1	7	25	5 $\beta$ -Pregnane-3,20-dione	128-23-4	47	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	54	1.35
[287]	Horse	ADH1	7	25	cyclohexanone	108-94-1	14100	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{ACT SITE}}^{-1}$	354	8.85
[288]	Human	ADH1	10	25	ethanol	64-17-5	1700		240	3.00
[288]	Human	ADH1	10	25	methanol	67-56-1	150000		12	0.15
[288]	Human	ADH1	10	25	ethylene glycol	107-21-1	50000		72	0.90
[288]	Human	ADH1	10	25	benzyl alcohol	100-51-6	12		280	3.50
[288]	Human	ADH1	10	25	octanol	111-87-5	8		260	3.25
[288]	Human	ADH1	10	25	16-hydroxyhexadecanoic acid	506-13-8	10		120	1.50
[288]	Human	ADH1	10	25	cyclohexanol	108-93-0	8		180	2.25
[288]	Human	ADH1	10	25	ethanol	64-17-5	1600		142	1.78
[288]	Human	ADH1	10	25	benzyl alcohol	100-51-6	4		122	1.53
[288]	Human	ADH1	10	25	cyclohexanol	108-93-0	6		118	1.48
[288]	Human	ADH1	10	25	ethanol	64-17-5	2000		200	2.50
[288]	Human	ADH1	10	25	methanol	67-56-1	15000		12	0.15
[288]	Human	ADH1	10	25	ethylene glycol	107-21-1	47000		50	0.63
[288]	Human	ADH1	10	25	benzyl alcohol	100-51-6	4		170	2.13
[288]	Human	ADH1	10	25	octanol	111-87-5	17		170	2.13
[288]	Human	ADH1	10	25	16-hydroxyhexadecanoic acid	506-13-8	9		100	1.25

[288]	Human	ADH1	10	25	cyclohexanol	108-93-0	5	140	1.75
[288]	Human	ADH1	10	25	ethanol	64-17-5	1000	80	1.00
[288]	Human	ADH1	10	25	methanol	67-56-1	30000	14	0.18
[288]	Human	ADH1	10	25	ethylene glycol	107-21-1	32000	66	0.83
[288]	Human	ADH1	10	25	benzyl alcohol	100-51-6	8	50	0.63
[288]	Human	ADH1	10	25	cyclohexanol	108-93-0	24	30	0.38
[288]	Human	ADH1	10	25	ethanol	64-17-5	1400	120	1.50
[288]	Human	ADH1	10	25	methanol	67-56-1	18000	12	0.15
[288]	Human	ADH1	10	25	ethylene glycol	107-21-1	36000	55	0.69
[288]	Human	ADH1	10	25	benzyl alcohol	100-51-6	11	58	0.73
[288]	Human	ADH1	10	25	octanol	111-87-5	5	70	0.88
[288]	Human	ADH1	10	25	16-hydroxyhexadecanoic acid	506-13-8	7	62	0.78
[288]	Human	ADH1	10	25	cyclohexanol	108-93-0	150	62	0.78
[288]	Human	ADH1	10	25	ethanol	64-17-5	2000	230	2.88
[288]	Human	ADH1	10	25	methanol	67-56-1	74000	22	0.28
[288]	Human	ADH1	10	25	ethylene glycol	107-21-1	57000	85	1.06
[288]	Human	ADH1	10	25	benzyl alcohol	100-51-6	7	127	1.59
[288]	Human	ADH1	10	25	16-hydroxyhexadecanoic acid	506-13-8	26	208	2.60
[288]	Human	ADH1	10	25	cyclohexanol	108-93-0	8	88	1.10
[288]	Human	ADH1	10	25	ethanol	64-17-5	1100	150	1.88
[288]	Human	ADH1	10	25	methanol	67-56-1	21000	12	0.15
[288]	Human	ADH1	10	25	ethylene glycol	107-21-1	220000	50	0.63
[288]	Human	ADH1	10	25	benzyl alcohol	100-51-6	26	115	1.44
[288]	Human	ADH1	10	25	octanol	111-87-5	9	52	0.65
[288]	Human	ADH1	10	25	16-hydroxyhexadecanoic acid	506-13-8	13	72	0.90

[288]	Human	ADH1	10	25	cyclohexanol	108-93-0	41			41		0.51
[288]	Human	ADH1	10	25	ethanol	64-17-5	1200			10		0.13
[288]	Human	ADH1	10	25	methanol	67-56-1	6000			8		0.10
[288]	Human	ADH1	10	25	ethylene glycol	107-21-1	13000			7		0.09
[288]	Human	ADH1	10	25	benzyl alcohol	100-51-6	120			11		0.14
[288]	Human	ADH1	10	25	16-hydroxyhexadecanoic acid	506-13-8	12			10		0.13
[288]	Human	ADH1	10	25	cyclohexanol	108-93-0	23000			8		0.10
[289]	Human	ADH3	10	25	pentanol	71-41-0	40000			14.7		0.18
[289]	Human	ADH3	10	25	octanol	111-87-5	1200			440		5.50
[289]	Human	ADH3	10	25	12-hydroxydodecanoic acid	505-95-3	100			182		2.28
[289]	Human	ADH3	10	25	vanillyl alcohol	498-00-0	11000			28		0.35
[289]	Human	ADH3	6.8	25	octanal	124-13-0	2400			200		2.50

ALDH

Ref.	Species	Isoenz	pH	T	Compound name	CAS	K <sub>m</sub> , μM	V <sub>max</sub>	V <sub>max</sub> units	k <sub>cat</sub> , min <sup>-1</sup>	V <sub>max</sub> , μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>
[290]	Human	ALDH1	7.1	25	3,4-dihydroxyphenyl acetaldehyde	5707-55-1	0.4	0.200	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.20
[290]	Human	ALDH1	7	25	4-aminobutanal	4390-05-0	760	0.130	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.13
[290]	Human	ALDH1	7	25	5-hydroxyindoleacetaldehyde	1892-21-3	2.4	0.220	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.22
[290]	Human	ALDH1	7	25	5-imidazoleacetaldehyde	na	39	0.180	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.18
[290]	Human	ALDH1	7	25	formaldehyde	50-00-0	330	0.460	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.46
[290]	Human	ALDH1	7	25	acetaldehyde	75-07-0	50	0.250	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.25
[290]	Human	ALDH1	7	25	propanal	123-38-6	5	0.280	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.28
[290]	Human	ALDH1	7	25	butanal	123-72-8	4	0.340	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.34

[290]	Human	ALDH1	7	25	pentanal	110-62-3	0.5	0.290	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.29
[290]	Human	ALDH1	7	25	hexanal	66-25-1	0.5	0.390	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.39
[290]	Human	ALDH2	7.1	25	3,4-dihydroxyphenyl acetaldehyde	5707-55-1	1	0.300	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.30
[290]	Human	ALDH2	7	25	4-aminobutanal	4390-05-0	512	0.150	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.15
[290]	Human	ALDH2	7	25	5-hydroxyindoleacetaldehyde	1892-21-3	0.8	0.210	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.21
[290]	Human	ALDH2	7	25	5-imidazoleacetaldehyde	na	30	0.280	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.28
[290]	Human	ALDH2	7	25	2-propenal	107-02-8	1.5	0.230	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.23
[290]	Human	ALDH2	7	25	acetaldehyde	75-07-0	1	0.310	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.31
[290]	Human	ALDH2	7	25	propanal	123-38-6	0.8	0.300	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.30
[290]	Human	ALDH2	7	25	butanal	123-72-8	0.5	0.250	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.25
[290]	Human	ALDH2	7	25	pentanal	110-62-3	0.5	0.370	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.37
[290]	Human	ALDH2	7	25	hexanal	66-25-1	0.5	0.500	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.50
[290]	Human	ALDH3	7	25	4-aminobutanal	4390-05-0	4.6	1.530	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.53
[290]	Human	ALDH3	7	25	3,4-dihydroxyphenyl acetaldehyde	5707-55-1	2.6	0.190	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.19
[290]	Human	ALDH3	7	25	5-imidazoleacetaldehyde	na	59	0.920	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.92
[290]	Human	ALDH3	7	25	2-propenal	107-02-8	4.9	0.380	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.38
[290]	Human	ALDH3	7	25	formaldehyde	50-00-0	410	0.190	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.19
[290]	Human	ALDH3	7	25	acetaldehyde	75-07-0	57	0.290	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.29
[290]	Human	ALDH3	7	25	propanal	123-38-6	9.5	0.300	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.30
[290]	Human	ALDH3	7	25	butanal	123-72-8	2.8	0.320	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.32
[290]	Human	ALDH3	7	25	pentanal	110-62-3	1.4	0.370	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.37
[290]	Human	ALDH3	7	25	hexanal	66-25-1	0.6	0.390	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.39
[291]	Human	ALDH3	7.4	25	betaine aldehyde	7418-61-3	260	6.500	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	6.50
[292]	Horse	ALDH1	7	25	isopentanal	590-86-3	0.5	1.1	relative to	0.14

[292]	Horse	ALDH1	7	25	pentanal	110-62-3	0.1	1.1	acetaldehyde relative to acetaldehyde	0.14
[292]	Horse	ALDH1	7	25	propanal	123-38-6	5	1.2	relative to acetaldehyde	0.16
[292]	Horse	ALDH1	7	25	acetaldehyde	75-07-0	70	1	relative to acetaldehyde	0.13
[292]	Horse	ALDH1	7	25	phenylacetaldehyde	122-78-1	4	1.9	relative to acetaldehyde	0.25
[292]	Horse	ALDH1	7	25	formaldehyde	50-00-0	940	0.8	relative to acetaldehyde	0.10
[292]	Horse	ALDH1	7	25	glycolaldehyde	141-46-8	130	0.9	relative to acetaldehyde	0.12
[292]	Horse	ALDH1	7	25	monochloroacetaldehyde	107-20-0	30	0.7	relative to acetaldehyde	0.09
[292]	Horse	ALDH1	7	25	benzaldehyde	100-52-7	0.1	0.6	relative to acetaldehyde	0.08
[292]	Horse	ALDH1	7	25	cinnamaldehyde	104-55-2	0.1	0.5	relative to acetaldehyde	0.07
[292]	Horse	ALDH2	7	25	isopentanal	590-86-3	0.1	0.6	relative to acetaldehyde	0.21
[292]	Horse	ALDH2	7	25	propanal	123-38-6	0.1	0.7	relative to acetaldehyde	0.25
[292]	Horse	ALDH2	7	25	acetaldehyde	75-07-0	0.2	1	relative to acetaldehyde	0.35
[292]	Horse	ALDH2	7	25	phenylacetaldehyde	122-78-1	0.3	1.4	relative to acetaldehyde	0.49
[292]	Horse	ALDH2	7	25	formaldehyde	50-00-0	270	1.6	relative to acetaldehyde	0.56
[292]	Horse	ALDH2	7	25	glycolaldehyde	141-46-8	50	3.2	relative to acetaldehyde	1.12
[292]	Horse	ALDH2	7	25	monochloroacetaldehyde	107-20-0	10	3.5	relative to acetaldehyde	1.23
[292]	Horse	ALDH2	7	25	benzaldehyde	100-52-7	0.1	0.11	relative to acetaldehyde	0.04



[292]	Horse	ALDH2	7	25	cinnamaldehyde	104-55-2	0.1	0.1	relative to acetaldehyde	0.04
[293]	Rat	ALDH2	7.4	25	propanal	123-38-6	0.08		26	0.12
[293]	Rat	ALDH2	9	25	propanal	123-38-6	0.1		68	0.31
[293]	Human	ALDH2	7.4	25	propanal	123-38-6	0.7		55	0.25
[294]	Human	ALDH1	9.5		acetaldehyde	75-07-0	120	0.800	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.80
[294]	Human	ALDH1	9.5		propanal	123-38-6	11	0.900	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.90
[294]	Human	ALDH2	9.5		acetaldehyde	75-07-0	2.4	1.700	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.70
[294]	Human	ALDH2	9.5		propanal	123-38-6	1.2	1.200	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.20
[295]	Rat	ALDH1	8	25	acetaldehyde	75-07-0	10000	10.000	$\mu\text{mol min}^{-1} \text{ml}^{-1}$	1.11
[295]	Rat	ALDH1	8	25	propanal	123-38-6	2000	12.000	$\mu\text{mol min}^{-1} \text{ml}^{-1}$	1.33
[295]	Rat	ALDH1	8	25	glutaraldehyde	111-30-8	560	8.000	$\mu\text{mol min}^{-1} \text{ml}^{-1}$	0.89
[295]	Rat	ALDH1	8	25	p-nitrobenzaldehyde	555-16-8	56	6.000	$\mu\text{mol min}^{-1} \text{ml}^{-1}$	0.67
[295]	Rat	ALDH1	8	25	(3,4-dihydroxyphenyl)(hydroxy) acetaldehyde	13023-73-9	23	0.900	$\mu\text{mol min}^{-1} \text{ml}^{-1}$	0.10
[295]	Rat	ALDH1	8	25	3,4-dihydroxyphenyl acetaldehyde	5707-55-1	17	0.830	$\mu\text{mol min}^{-1} \text{ml}^{-1}$	0.09
[295]	Rat	ALDH1	8	25	5-hydroxyindoleacetaldehyde	1892-21-3	26	0.820	$\mu\text{mol min}^{-1} \text{ml}^{-1}$	0.09
[295]	Rat	ALDH1	8	25	acetaldehyde	75-07-0	220	1.200	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.20
[295]	Rat	ALDH1	8	25	propanal	123-38-6	500	1.400	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.40
[295]	Rat	ALDH1	8	25	glutaraldehyde	111-30-8	50	2.000	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	2.00
[295]	Rat	ALDH1	8	25	p-nitrobenzaldehyde	555-16-8	2000	2.000	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	2.00
[295]	Rat	ALDH2	8	25	acetaldehyde	75-07-0	5	9.000	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	9.00
[295]	Rat	ALDH2	8	25	propanal	123-38-6	3.3	2.000	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	2.00
[295]	Rat	ALDH2	8	25	glutaraldehyde	111-30-8	6.2	4.500	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	4.50
[295]	Rat	ALDH2	8	25	p-nitrobenzaldehyde	555-16-8	18	1.800	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.80

[296]	Human	ALDH1	7.4	25	methylglyoxal	78-98-8	48	0.067	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.07
[296]	Human	ALDH1	7.4		acetaldehyde	75-07-0	30	0.280	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.28
[296]	Human	ALDH1	9	25	methylglyoxal	78-98-8	24	0.120	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.12
[296]	Human	ALDH1	9		acetaldehyde	75-07-0	40	0.800	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.80
[296]	Human	ALDH1	7.4	25	glycolaldehyde	141-46-8	243	0.340	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.34
[296]	Human	ALDH2	7.4	25	methylglyoxal	78-98-8	8.6	0.060	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.06
[296]	Human	ALDH2	7.4		acetaldehyde	75-07-0	3	0.400	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.40
[296]	Human	ALDH2	9	25	methylglyoxal	78-98-8	21	0.290	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.29
[296]	Human	ALDH2	9		acetaldehyde	75-07-0	2	1.700	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.70
[296]	Human	ALDH2	7.4	25	glycolaldehyde	141-46-8	46	2.000	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	2.00
[296]	Human	ALDH3	7.4	25	methylglyoxal	78-98-8	586	1.100	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.10
[296]	Human	ALDH3	7.4	25	methylglyoxal	78-98-8	552	0.800	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.80
[296]	Human	ALDH3	9	25	methylglyoxal	78-98-8	1876	4.000	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	4.00
[296]	Human	ALDH3	9	25	methylglyoxal	78-98-8	958	3.400	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	3.40
[296]	Human	ALDH3	7.4	25	betaine aldehyde	7418-61-3	90	2.800	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	2.80
[80]	Human	ALDH1	7.4	25	4-aminobutanol	4390-05-0	5			na
[80]	Human	ALDH1	7.4	25	acetaldehyde	75-07-0	50			na
[80]	Human	ALDH1	7.4	25	glycolaldehyde	141-46-8	240			na
[80]	Human	ALDH1	7.4	25	betaine aldehyde	7418-61-3	260			na
[80]	Human	ALDH1	7.4	25	N-acetyl-4-aminobutanol	na	100			na
[70]	Human	ALDH1	9.5	25	acetaldehyde	75-07-0	180		790	3.43
[70]	Human	ALDH1	9.5	25	propanal	123-38-6	4.5		700	3.04
[70]	Human	ALDH1	9.5	25	pentanal	110-62-3	0.16		490	2.13
[70]	Human	ALDH1	9.5	25	hexanal	66-25-1	0.041		250	1.09
[70]	Human	ALDH1	9.5	25	heptanal	111-71-7	0.018		260	1.13

[70]	Human	ALDH1	9.5	25	octanal	124-13-0	0.012	250	1.09
[70]	Human	ALDH1	9.5	25	decanal	112-81-2	0.0029	230	1.00
[70]	Human	ALDH1		25	5-bromo-1-naphthaldehyde	na	0.0025	9.5	0.04
[70]	Human	ALDH1		25	6-(dimethylamino)-2-naphthaldehyde	na	0.0063	160	0.70
[70]	Human	ALDH1		25	5-nitro-1-naphthaldehyde	na	0.011	33	0.14
[70]	Human	ALDH1		25	fluorene-2-carboxaldehyde	30084-90-3	0.054	360	1.57
[70]	Human	ALDH1		25	p-(dimethylamino)-benzaldehyde	100-10-7	0.06	128	0.56
[70]	Human	ALDH1		25	4-methoxy-1-naphthaldehyde	15971-29-6	0.2	265	1.15
[70]	Human	ALDH1		25	indole-3-acetaldehyde	2591-98-2	0.31	680	2.96
[70]	Human	ALDH1		25	trans-cinnamaldehyde	14371-10-9	0.4	470	2.04
[70]	Human	ALDH1		25	p-(dimethylamino)cinnamaldehyde	6203-18-5	0.9	490	2.13
[70]	Human	ALDH1		25	7-(dimethylamino)-coumarin-4-carboxaldehyde	na	1.42	474	2.06
[70]	Human	ALDH1		25	phenylacetaldehyde	122-78-1	5.5	3000	13.04
[70]	Human	ALDH2	9.5	25	formaldehyde	50-00-0	320	4050	16.88
[70]	Human	ALDH2	9.5	25	acetaldehyde	75-07-0	0.2	1180	4.92
[70]	Human	ALDH2	9.5	25	propanal	123-38-6	0.095	1180	4.92
[70]	Human	ALDH2	9.5	25	pentanal	110-62-3	0.034	1370	5.71
[70]	Human	ALDH2	9.5	25	hexanal	66-25-1	0.03	1710	7.13
[70]	Human	ALDH2	9.5	25	heptanal	111-71-7	0.027	1360	5.67
[70]	Human	ALDH2	9.5	25	octanal	124-13-0	0.028	900	3.75
[70]	Human	ALDH2	9.5	25	decanal	112-81-2	0.022	700	2.92
[70]	Human	ALDH2		25	p-nitrocinnamaldehyde	1734-79-8	0.0007	27	0.11

[70]	Human	ALDH2	p- de	25	(dimethylamino)cinnamaldehyde	6203-18-5	0.005	90	0.38
[70]	Human	ALDH2		25	trans-cinnamaldehyde	14371-10-9	0.035	150	0.63
[70]	Human	ALDH2		25	3-phenylpropanal	104-53-0	0.5	740	3.08
[70]	Human	ALDH2		25	2-phenylpropanal	93-53-8	0.93	910	3.79
[70]	Human	ALDH2		25	phenylacetaldehyde	122-78-1	0.029	1800	7.50
[70]	Human	ALDH2		25	2,4-dinitrobenzaldehyde	528-75-6	0.0032	40	0.17
[70]	Human	ALDH2		25	o-nitrobenzaldehyde	552-89-6	0.0063	74	0.31
[70]	Human	ALDH2		25	p-nitrobenzaldehyde	555-16-8	0.007	430	1.79
[70]	Human	ALDH2		25	benzaldehyde	100-52-7	0.018	350	1.46
[70]	Human	ALDH2		25	p-methylbenzaldehyde	104-87-0	0.017	180	0.75
[70]	Human	ALDH2		25	m-methylbenzaldehyde	620-23-5	0.018	270	1.13
[70]	Human	ALDH2		25	p-methoxybenzaldehyde	123-11-5	0.018	150	0.63
[70]	Human	ALDH2		25	p-(dimethylamino)-benzaldehyde	100-10-7	0.02	140	0.58
[70]	Human	ALDH2		25	m-methoxybenzaldehyde	591-31-1	0.09	350	1.46
[70]	Human	ALDH2		25	m-hydroxybenzaldehyde	100-83-4	0.24	60	0.25
[70]	Human	ALDH2		25	3,4-dimethoxybenzaldehyde	120-14-9	0.33	85	0.35
[70]	Human	ALDH2		25	o-methoxybenzaldehyde	135-02-4	0.8	22	0.09
[70]	Human	ALDH2		25	o-methylbenzaldehyde	529-20-4	1.3	165	0.69
[70]	Human	ALDH2		25	o-hydroxybenzaldehyde	90-02-8	320	590	2.46
[70]	Human	ALDH2		25	5-bromo-1-naphthaldehyde	na	0.0004	15	0.06
[70]	Human	ALDH2		25	5-nitro-1-naphthaldehyde	na	0.0004	50	0.21
[70]	Human	ALDH2		25	6-[O-(CH <sub>2</sub> ) <sub>5</sub> -COOH]-2-naphthaldehyde	na	0.0009	28	0.12
[70]	Human	ALDH2		25	6-(dimethylamino)-2-	na	0.0023	37	0.15

[70]	Human	ALDH2	25	naphthaldehyde	66-99-9	0.008	240	1.00
[70] <sup>3</sup>	Human	ALDH2	25	4-methoxy-1-naphthaldehyde	15971-29-6	0.065	<0.7	0.00
[70]	Human	ALDH2	25	7-acetoxy-coumarin-4-carboxaldehyde	na	0.06	65	0.27
[70]	Human	ALDH2	25	7-(dimethylamino)-coumarin-4-carboxaldehyde	na	0.062	78	0.33
[70]	Human	ALDH2	25	7-methoxy-coumarin-4-carboxaldehyde	na	0.28	190	0.79
[70]	Human	ALDH2	25	6,7-dimethoxy-coumarin-4-carboxaldehyde	na	0.69	59	0.25
[70]	Human	ALDH2	25	7-hydroxy-coumarin-4-carboxaldehyde	na	150	500	2.08
[70]	Human	ALDH2	25	quinoline-3-carboxaldehyde	13669-42-6	0.33	670	2.79
[70]	Human	ALDH2	25	7-(dimethylamino)-2-quinolinone-4-carboxaldehyde	na	2.13	130	0.54
[70]	Human	ALDH2	25	quinoline-4-carboxaldehyde	4363-93-3	2.8	280	1.17
[70]	Human	ALDH2	25	6-methoxy-2-quinolinone-4-carboxaldehyde	na	5.4	59	0.25
[70]	Human	ALDH2	25	7-methoxy-2-quinolinone-4-carboxaldehyde	na	50	320	1.33
[70]	Human	ALDH2	25	phenanthrene-9-carboxaldehyde	4707-71-5	0.004	18	0.08
[70] <sup>4</sup>	Human	ALDH2	25	indole-3-aldehyde	487-89-8	10	<2.4	0.01
[70]	Human	ALDH2	25	indole-3-acetaldehyde	2591-98-2	0.15	560	2.33
[70] <sup>5</sup>	Human	ALDH2	25	5-methoxyindole-3-carboxaldehyde	10601-19-1	18.9	<2	0.01

<sup>3</sup> Data not inserted as below the detection limit.

<sup>4</sup> Data not inserted as below the detection limit.

<sup>5</sup> Data not inserted as below the detection limit.

[70]	Human	ALDH2	25	3-pyridinaldehyde	500-22-1	1.96			1660	6.92
[297]	Human	ALDH3	7.4	4-aminobutanal	4390-05-0	13.8	2.000	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		2.00
[297]	Human	ALDH3	7.4	4-aminobutanal	4390-05-0	8	1.400	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		1.40
[297]	Human	ALDH3	7.4	propanal	123-38-6	8	0.310	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.31
[297]	Human	ALDH3	7.4	propanal	123-38-6	9.6	0.200	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.20
[297]	Human	ALDH3	7.4	acetaldehyde	75-07-0	50.4	0.280	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.28
[297]	Human	ALDH3	7.4	acetaldehyde	75-07-0	40.3	0.200	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.20
[298]	Rat	ALDH 1	8.5	propanal	123-38-6	150	1	relative to propionaldehyde		4.35
[298]	Rat	ALDH 1	8.5	benzaldehyde	100-52-7	1600	0.35	relative to propionaldehyde		1.52
[298]	Rat	ALDH 1	8.5	acetaldehyde	75-07-0	2400	0.09	relative to propionaldehyde		0.41
[298]	Rat	ALDH 1	8.5	benzaldehyde	100-52-7	5900	1.4	relative to propionaldehyde		6.44
[298]	Rat	ALDH 2	8.5	propanal	123-38-6	150	1	relative to propionaldehyde		9.93
[298]	Rat	ALDH 2	8.5	benzaldehyde	100-52-7	50000	0.8	relative to propionaldehyde		7.94
[298]	Rat	ALDH 2	8.5	propanal	123-38-6	1700	1	relative to propionaldehyde		5.69
[298]	Rat	ALDH 2	8.5	benzaldehyde	100-52-7	3000	0.34	relative to propionaldehyde		1.93
[299]	Rat	ALDH	7.4	decanal	112-81-2	7.1	0.330	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.33
[300]	Human	ALDH1	7.5	acetaldehyde	75-07-0	180			380	1.55
[300]	Human	ALDH1	9.5	acetaldehyde	75-07-0	180			790	3.22
[300]	Human	ALDH2	7.5	acetaldehyde	75-07-0	0.2			280	1.24
[300]	Human	ALDH2	9.5	acetaldehyde	75-07-0	0.2			1180	5.24
[301]	Rat	ALDH1	9.6	benzaldehyde	100-52-7	20	0.009	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.01
[301]	Rat	ALDH1	9.6	o-fluorobenzaldehyde	446-52-6	8	0.033	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.03

[301]	Rat	ALDH1	9.6	37	o-chlorobenzaldehyde	89-98-5	6	0.029	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.03
[301]	Rat	ALDH1	9.6	37	o-bromobenzaldehyde	6630-33-7	12	0.014	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.01
[301]	Rat	ALDH1	9.6	37	p-fluorobenzaldehyde	459-57-4	12	0.008	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.01
[301]	Rat	ALDH1	9.6	37	p-chlorobenzaldehyde	104-88-1	31	0.021	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.02
[301]	Rat	ALDH1	9.6	37	p-bromobenzaldehyde	1122-91-4	39	0.025	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.02
[301]	Rat	ALDH1	9.6	37	p-iodobenzaldehyde	15164-44-0	81	0.036	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.04
[301]	Rat	ALDH2	9.6	37	benzaldehyde	100-52-7	1805	0.039	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.04
[301]	Rat	ALDH2	9.6	37	o-fluorobenzaldehyde	446-52-6	703	0.019	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.02
[301]	Rat	ALDH2	9.6	37	o-bromobenzaldehyde	6630-33-7	379	0.001	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.00
[301]	Rat	ALDH2	9.6	37	p-fluorobenzaldehyde	459-57-4	2057	0.040	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.04
[301]	Rat	ALDH2	9.6	37	p-chlorobenzaldehyde	104-88-1	826	0.031	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.03
[301]	Rat	ALDH2	9.6	37	p-bromobenzaldehyde	1122-91-4	616	0.024	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.02
[301]	Rat	ALDH2	9.6	37	p-iodobenzaldehyde	15164-44-0	1103	0.013	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.01
[301]	Rat	ALDH3	9.6	37	benzaldehyde	100-52-7	733	0.045	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.05
[301]	Rat	ALDH3	9.6	37	o-fluorobenzaldehyde	446-52-6	822	0.020	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.02
[301]	Rat	ALDH3	9.6	37	o-chlorobenzaldehyde	89-98-5	63	0.002	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.00
[301]	Rat	ALDH3	9.6	37	o-bromobenzaldehyde	6630-33-7	67	0.002	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.00
[301]	Rat	ALDH3	9.6	37	p-fluorobenzaldehyde	459-57-4	859	0.056	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.06
[301]	Rat	ALDH3	9.6	37	p-chlorobenzaldehyde	104-88-1	1159	0.139	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.14
[301]	Rat	ALDH3	9.6	37	p-bromobenzaldehyde	1122-91-4	617	0.112	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.11
[301]	Rat	ALDH3	9.6	37	p-iodobenzaldehyde	15164-44-0	241	0.082	$\mu\text{mol}_{\text{NADH}} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.08
[302]	Rat	ALDH1	7.4	25	formaldehyde	50-00-0	178	1.360	$\mu\text{mol} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.36
[302]	Rat	ALDH1	7.4	25	acetaldehyde	75-07-0	0.4	0.561	$\mu\text{mol} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.56
[302]	Rat	ALDH1	7.4	25	propanal	123-38-6	0.7	0.495	$\mu\text{mol} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.50
[302]	Rat	ALDH1	7.4	25	butanal	123-72-8	0.7	0.831	$\mu\text{mol} \text{min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.83

[302]	Rat	ALDH1	7.4	25	hexanal	66-25-1	0.8	1.220	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.22
[302]	Rat	ALDH1	7.4	25	p- de (dimethylamino)cinnamaldehy	6203-18-5	0.1	0.120	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.12
[302]	Rat	ALDH1	7.4	25	benzaldehyde	100-52-7	0.1	0.072	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.07
[302]	Rat	ALDH1	7.4	25	p-nitrobenzaldehyde	555-16-8	0.1	0.275	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.28
[302]	Rat	ALDH1	7.4	25	p-methoxybenzaldehyde	123-11-5	0.1	0.015	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.02
[302]	Rat	ALDH1	7.4	25	glutaraldehyde	111-30-8	0.9	0.566	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.57
[302]	Rat	ALDH1	7.4	25	glycolaldehyde	141-46-8	38	1.690	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.69
[303]	Rat	ALDH 1	9.6		formaldehyde	50-00-0	31	1.2	relative to propionaldehyde	0.28
[303]	Rat	ALDH 1	9.6		acetaldehyde	75-07-0	1.5	0.97	relative to propionaldehyde	0.22
[303]	Rat	ALDH 1	9.6		propanal	123-38-6	31	1	relative to propionaldehyde	0.23
[303]	Rat	ALDH 1	9.6		butanal	123-72-8	46	0.83	relative to propionaldehyde	0.19
[303]	Rat	ALDH 1	9.6		isobutanal	78-84-2	36	0.59	relative to propionaldehyde	0.14
[303]	Rat	ALDH 1	9.6		pentanal	110-62-3	0.7	0.007	relative to propionaldehyde	0.00
[303]	Rat	ALDH 1	9.6		heptanal	111-71-7	5.6	0.72	relative to propionaldehyde	0.17
[303]	Rat	ALDH 1	9.6		glycolaldehyde	141-46-8	45	2.94	relative to propionaldehyde	0.68
[303]	Rat	ALDH 1	9.6		monochloroacetaldehyde	107-20-0	43	3.17	relative to propionaldehyde	0.73
[303]	Rat	ALDH 1	9.6		glyoxal	107-22-2	43	0.29	relative to propionaldehyde	0.07
[303]	Rat	ALDH 1	9.6		2-butenal	4170-30-3	32	0.12	relative to propionaldehyde	0.03
[303]	Rat	ALDH 1	9.6		methylglyoxal	78-98-8	13	0.33	relative to propionaldehyde	0.08



[303]	Rat	ALDH 1	9.6	p-carboxybenzaldehyde	619-66-9	24	0.23	relative to propionaldehyde	0.05
[303]	Rat	ALDH 1	9.6	p-cyanobenzaldehyde	105-07-7	14	0.62	relative to propionaldehyde	0.14
[303]	Rat	ALDH 1	9.6	phenylacetaldehyde	122-78-1	27	0.88	relative to propionaldehyde	0.20
[303]	Rat	ALDH 2	9.6	formaldehyde	50-00-0	nd	nd	relative to propionaldehyde	nd
[303]	Rat	ALDH 2	9.6	acetaldehyde	75-07-0	1500	0.67	relative to propionaldehyde	0.53
[303]	Rat	ALDH 2	9.6	propanal	123-38-6	450	1	relative to propionaldehyde	0.80
[303]	Rat	ALDH 2	9.6	butanal	123-72-8	740	0.61	relative to propionaldehyde	0.49
[303]	Rat	ALDH 2	9.6	isobutanal	78-84-2	380	0.3	relative to propionaldehyde	0.24
[303]	Rat	ALDH 2	9.6	pentanal	110-62-3	23	0.78	relative to propionaldehyde	0.62
[303]	Rat	ALDH 2	9.6	heptanal	111-71-7	47	0.41	relative to propionaldehyde	0.33
[303]	Rat	ALDH 2	9.6	glycolaldehyde	141-46-8	160	0.3	relative to propionaldehyde	0.24
[303]	Rat	ALDH 2	9.6	monochloroacetaldehyde	107-20-0	270	1.96	relative to propionaldehyde	1.56
[303]	Rat	ALDH 2	9.6	glyoxal	107-22-2	1900	0.45	relative to propionaldehyde	0.36
[303]	Rat	ALDH 2	9.6	2-butenal	4170-30-3	1900	0.3	relative to propionaldehyde	0.24
[303]	Rat	ALDH 2	9.6	methylglyoxal	78-98-8	3200	0.62	relative to propionaldehyde	0.49
[303]	Rat	ALDH 2	9.6	p-carboxybenzaldehyde	619-66-9	41	0.4	relative to propionaldehyde	0.32
[303]	Rat	ALDH 2	9.6	p-cyanobenzaldehyde	105-07-7	5	1.32	relative to propionaldehyde	1.05
[303]	Rat	ALDH 2	9.6	phenylacetaldehyde	122-78-1	4	0.74	relative to propionaldehyde	0.59

FMO

Ref.	Species	Isoenz	pH	T	Compound name	CAS	K <sub>m</sub> , μM	V <sub>max</sub>	V <sub>max</sub> units	k <sub>cat</sub> , min <sup>-1</sup>	V <sub>max</sub> μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>
[304]	Pig	FMO	7.4	37	MPTP (1 -methyl-4-phenyl-1,2,3,6-tetrahydropyridine)	28289-54-5	32.0	730	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.73
[304]	Pig	FMO	7.4	37	amitriptyline	50-48-6	98.0	740	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.74
[304]	Pig	FMO	7.4	37	imipramine	50-49-7	22.0	730	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.73
[304]	Pig	FMO	7.4	37	pargyline	555-57-7	65.0	750	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.75
[304]	Pig	FMO	7.4	37	selegiline	14611-51-9	49.0	750	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.75
[304]	Pig	FMO	7.4	37	clorgyline	17780-72-2	2.0	730	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.73
[305]	Pig	FMO	7.5	37	4-tolyl ethyl sulfide	622-63-9	13			48	0.86
[305]	Pig	FMO	7.5	37	thioanisole	100-68-5	16.5			62	1.11
[305]	Pig	FMO	7.5	37	benzyl methyl sulfide	766-92-7	1.5			49	0.88
[305]	Pig	FMO	7.5	37	sulindac sulfide	32004-67-4	3.0			80	1.43
[306]	Pig	FMO			guanethidine	55-65-2	310	0.56	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>		0.56
[81]	Pig	FMO	7.5	37	N,N-dimethyl-2-[2-(trifluoromethyl)-10H-phenothiazin-10-yl]ethanamine	na	55.0		μmol min <sup>-1</sup> μmol <sub>FMO</sub> <sup>-1</sup>	56	1.00
[81]	Pig	FMO	7.5	37	triflupromazine	146-54-3	11.0		μmol min <sup>-1</sup> μmol <sub>FMO</sub> <sup>-1</sup>	59	1.05
[81]	Pig	FMO	7.5	37	N,N-dimethyl-4-[2-(trifluoromethyl)-10H-phenothiazin-10-yl]butan-1-amine	na	11.0		μmol min <sup>-1</sup> μmol <sub>FMO</sub> <sup>-1</sup>	57	1.02
[81]	Pig	FMO	7.5	37	N,N-dimethyl-5-[2-(trifluoromethyl)-10H-phenothiazin-10-yl]pentan-1-amine	na	11.0		μmol min <sup>-1</sup> μmol <sub>FMO</sub> <sup>-1</sup>	60	1.07

[81]	Pig	FMO	7.5	37	N,N-dimethyl-6-[2-(trifluoromethyl)-10H-phenothiazin-10-yl]esan-1-amine	na	14.0	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{FMO}}^{-1}$	67	1.20
[81]	Pig	FMO	7.5	37	N,N-dimethyl-7-[2-(trifluoromethyl)-10H-phenothiazin-10-yl]heptan-1-amine	na	15.0	$\mu\text{mol min}^{-1} \mu\text{mol}_{\text{FMO}}^{-1}$	68	1.21
[81]	Pig	FMO	7.5	37	thiourea	62-56-6	23			na
[81]	Pig	FMO	7.5	37	phenylthiocarbamide	103-85-5	4			na
[81]	Pig	FMO	7.5	37	1-naphthylthiourea	86-88-4	4			na
[81]	Pig	FMO	7.5	37	thiocarbamide	102-08-9	7			na
[81]	Pig	FMO	7.5	37	phenothiazine	92-84-2	12			na
[81]	Pig	FMO	7.5	37	2-(trifluoromethyl)phenothiazine	92-30-8	500			na
[307]	Pig	FMO	7.4	37	ethyl methyl sulfide	624-89-5	1380.0	$78.74 \text{ nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.08
[307]	Pig	FMO	7.4	37	p-chlorophenyl methyl sulfide	123-09-1	185.0	$103.00 \text{ nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.10
[307]	Pig	FMO	7.4	37	diphenyl sulphide	139-66-2	68.0	$49.26 \text{ nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.05
[308]	Pig	FMO	8.3	37	cysteamine	60-23-1	120	$1.00 \mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		1.00
[308]	Pig	FMO	8.3	37	2-mercaptoethanol	60-24-2	2800	$0.90 \mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.90
[308]	Pig	FMO	8.3	37	thioacetamide	62-55-5	65	$0.90 \mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.90
[308]	Pig	FMO	8.3	37	thiourea	62-56-6	41	$1.00 \mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		1.00
[308]	Pig	FMO	8.3	37	2-mercaptobenzimidazole	583-39-1	37	$1.10 \mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		1.10
[308]	Pig	FMO	8.3	37	phenylisothiocyanate	103-72-0	3300	$0.80 \mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.80
[82]	Pig	FMO	7.4		2-naphthylamine	91-59-8	1200			na
[82]	Pig	FMO	7.4		2-aminoazulene	na	150			na
[82]	Pig	FMO	7.4		rosaniline	632-99-5	175			na
[82]	Pig	FMO	7.4		auramine	492-80-8	19			na

[82]	Pig	FMO	7.4	trimethylamine	75-50-3	3200	na
[82]	Pig	FMO	7.4	butanethiol	109-79-5	33	na
[82]	Pig	FMO	7.4	2-butanethiol	513-53-1	28	na
[82]	Pig	FMO	7.4	tert-butylthiol	75-66-1	50	na
[82]	Pig	FMO	7.4	isobutanethiol	513-44-0	53	na
[82]	Pig	FMO	7.4	1-hexanethiol	111-31-9	15	na
[82]	Pig	FMO	7.4	1-heptanethiol	1639-09-4	15	na
[82]	Pig	FMO	7.4	1,4-butanedithiol	1191-08-8	3	na
[82]	Pig	FMO	7.4	butyldisulfide	na	6	na
[82]	Pig	FMO	7.4	benzylsulfide	3492-66-8	45	na
[82]	Pig	FMO	7.4	methylsulfide	74-93-1	8	na
[82]	Pig	FMO	7.4	ethylene sulfide	420-12-2	60	na
[82]	Pig	FMO	7.4	thionazane	50-52-2	90	na
[82]	Pig	FMO	8.4	4-chloro-N-methylaniline	932-96-7	430	na
[82]	Pig	FMO	8.4	2-naphthylamine	91-59-8	1100	na
[82]	Pig	FMO	8.4	rosaniline	632-99-5	93	na
[82]	Pig	FMO	8.4	acetopromazine	61-00-7	53	na
[82]	Pig	FMO	8.4	trimeprazine	84-96-8	20	na
[82]	Pig	FMO	8.4	methotrimeprazine	60-99-1	90	na
[82]	Pig	FMO	8.4	diethazine	60-91-3	60	na
[82]	Pig	FMO	8.4	prothipendyl	303-69-5	130	na
[82]	Pig	FMO	8.4	butriptyline	15686-37-0	100	na
[82]	Pig	FMO	8.4	benzphetamine	156-08-1	128	na
[82]	Pig	FMO	8.4	methamphetamine	537-46-2	560	na
[309]	Pig	FMO	8.4	methimazole	60-56-0	13	890 nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>

[309]	Pig	FMO	8.4	38	1-naphthylthiourea	86-88-4	4	747	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.75
[309]	Pig	FMO	8.4	38	phenylthiocarbamide	103-85-5	3	811	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.81
[309]	Pig	FMO	8.4	38	thioglycerol	96-27-5	4900	721	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.72
[309]	Pig	FMO	8.4	38	dithiothreitol	3483-12-3	465	685	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.69
[309]	Pig	FMO	8.4	38	trans-o-dithiane-4,5-diol	14193-38-5	270	739	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.74
[83]	Pig	FMO	7.4	37	thiourea	62-56-6	23	0.54- 0.56	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	na
[83]	Pig	FMO	7.4	37	thiocarbamide	102-08-9	6.7	0.54- 0.56	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	na
[83]	Pig	FMO	7.4	37	2-mercaptobenzimidazole	583-39-1	13	0.54- 0.56	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	na
[310]	Pig	FMO	7.7	25	dimethylaniline	121-69-7	14	107.1	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.11
[310]	Pig	FMO	7.7	25	methylhydrazine	60-34-4	35000	154.0	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.15
[310]	Pig	FMO	7.7	25	1,1-dimethylhydrazine	57-14-7	430	86.9	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.09
[310]	Pig	FMO	7.7	25	1,2-dimethylhydrazine	540-73-8	5600	35.5	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.04
[310]	Pig	FMO	7.7	25	procabazine	671-16-9	1800	28.1	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.03
[311]	Pig	FMO	8.2	37	dimethylaniline	121-69-7	20	950	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.95
[311]	Pig	FMO	8.2	37	methylhydrazine	60-34-4	35000	1370	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.37
[311]	Pig	FMO	8.2	37	ethylhydrazine	624-80-6	40000	1350	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.35
[311]	Pig	FMO	8.2	37	n-propylhydrazine	5039-61-2	15000	1520	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.52
[311]	Pig	FMO	8.2	37	isopropylhydrazine	2257-52-5	8300	603	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.60
[311]	Pig	FMO	8.2	37	butylhydrazine	3530-11-8	6900	1480	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.48
[311]	Pig	FMO	8.2	37	phenylhydrazine	100-63-0	3000	890	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.89
[311]	Pig	FMO	8.2	37	benzylhydrazine	555-96-4	7000	880	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.88
[311]	Pig	FMO	8.2	37	1,2-dimethylhydrazine	540-73-8	12000	644	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.64
[311]	Pig	FMO	8.2	37	1-methyl-2-benzylhydrazine	10309-79-2	2000	1260	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.26

[311]	Pig	FMO	8.2	37	procarbazine	671-16-9	5700	560	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.56
[311]	Pig	FMO	8.2	37	1,1-dimethylhydrazine	57-14-7	430	890	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.89
[311]	Pig	FMO	8.2	37	1-methyl-1-phenylhydrazine	618-40-6	80	1130	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.13
[311]	Pig	FMO	8.2	37	1,2-dimethylphenylhydrazine	na	380	895	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.90
[311]	Pig	FMO	8.2	37	N-aminopyrrolidine	16596-41-1	100	960	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.96
[311]	Pig	FMO	8.2	37	N-aminomorpholine	4319-49-7	610	950	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.95
[311]	Pig	FMO	8.2	37	N-aminopiperidine	2213-43-6	30	960	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.96
[311]	Pig	FMO	8.2	37	N-aminomopiperidine	5906-35-4	170	878	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.88
[312]	Mouse	FMO	8.1		trimethylamine	75-50-3	2340	1.38	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.38
[312]	Mouse	FMO	8.1		triethylamine	121-44-8	2890	0.85	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.85
[312]	Mouse	FMO	8.1		N-methylaniline	100-61-8	1060	1.39	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.39
[312]	Mouse	FMO	8.1		dimethylaniline	121-69-7	105	1.45	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.45
[312]	Mouse	FMO	8.1		N,N-diethylaniline	91-66-7	144	1.39	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.39
[312]	Mouse	FMO	8.1		imipramine	50-49-7	27	1.48	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.48
[312]	Mouse	FMO	8.1		ethylmorphine	76-58-4	1650	0.63	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.63
[312]	Mouse	FMO	8.1		dimethyl sulfide	75-18-3	34	1.37	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.37
[312]	Mouse	FMO	8.1		thioanisole	100-68-5	6	1.46	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.46
[312]	Mouse	FMO	8.1		benzyl methyl sulfide	766-92-7	2	1.56	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.56
[312]	Mouse	FMO	8.1		trans-o-dithiane-4,5-diol	14193-38-5	1310	0.98	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.98
[312]	Mouse	FMO	8.1		cysteamine	60-23-1	65	1.49	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.49
[312]	Mouse	FMO	8.1		butanethiol	109-79-5	262	1.12	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.12
[312]	Mouse	FMO	8.1		benzyl mercaptan	100-53-8	330	1.38	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.38
[312]	Mouse	FMO	8.1		dithiothreitol	3483-12-3	847	0.73	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.73
[312]	Mouse	FMO	8.1		methimazole	60-56-0	9	1.05	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.05
[312]	Mouse	FMO	8.1		2-mercaptobenzimidazole	583-39-1	24	1.35	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.35

[312]	Mouse	FMO	8.1	thioacetamide	62-55-5	10	1.64	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.64
[312]	Mouse	FMO	8.1	thiourea	62-56-6	11	1.65	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.65
[312]	Mouse	FMO	8.1	phenylthiocarbamide	103-85-5	7	1.4	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.40
[312]	Mouse	FMO	8.1	thiocarbamide	102-08-9	66	1.05	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.05
[312]	Mouse	FMO	8.1	thiobenzamide	2227-79-4	8	1.42	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.42
[312]	Pig	FMO	8.1	trimethylamine	75-50-3	617	0.42	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.42
[312]	Pig	FMO	8.1	triethylamine	121-44-8	1090	0.33	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.33
[312]	Pig	FMO	8.1	N-methylaniline	100-61-8	343	0.47	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.47
[312]	Pig	FMO	8.1	dimethylaniline	121-69-7	11	0.5	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.50
[312]	Pig	FMO	8.1	N,N-diethylaniline	91-66-7	44	0.48	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.48
[312]	Pig	FMO	8.1	imipramine	50-49-7	7	0.46	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.46
[312]	Pig	FMO	8.1	ethylmorphine	76-58-4	284	0.35	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.35
[312]	Pig	FMO	8.1	dimethyl sulfide	75-18-3	11	0.39	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.39
[312]	Pig	FMO	8.1	thioanisole	100-68-5	2	0.33	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.33
[312]	Pig	FMO	8.1	benzyl methyl sulfide	766-92-7	2	0.38	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.38
[312]	Pig	FMO	8.1	trans-o-dithiane-4,5-diol	14193-38-5	254	0.35	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.35
[312]	Pig	FMO	8.1	cysteamine	60-23-1	59	0.46	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.46
[312]	Pig	FMO	8.1	butanethiol	109-79-5	78	0.33	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.33
[312]	Pig	FMO	8.1	benzyl mercaptan	100-53-8	108	0.36	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.36
[312]	Pig	FMO	8.1	dithiothreitol	3483-12-3	296	0.18	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.18
[312]	Pig	FMO	8.1	methimazole	60-56-0	13	0.42	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.42
[312]	Pig	FMO	8.1	2-mercaptobenzimidazole	583-39-1	16	0.4	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.40
[312]	Pig	FMO	8.1	thioacetamide	62-55-5	10	0.4	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.40
[312]	Pig	FMO	8.1	thiourea	62-56-6	20	0.45	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.45
[312]	Pig	FMO	8.1	phenylthiocarbamide	103-85-5	20	0.45	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.45

[312]	Pig	FMO	8.1	thiocarbamide	102-08-9	13	0.34	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.34
[312]	Pig	FMO	8.1	thiobenzamide	2227-79-4	3	0.34	$\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.34
[84]	Pig	FMO	7.6	fonofos	944-22-9	33		na	na
[84]	Pig	FMO	7.6	S-phenyl diethylphosphinothiolothionate	na	48		na	na
[84]	Pig	FMO	7.6	diethylphenylphosphine sulfide	na	99		na	na
[84]	Pig	FMO	7.6	diethylphenylphosphine	1605-53-4	2.5		na	na
[313]	Mouse	FMO	8.1	thiourea	62-56-6	19.9	1715.4	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.72
[313]	Mouse	FMO	8.1	disulfoton	298-04-4	3.4	1693.3	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.69
[313]	Mouse	FMO	8.1	demeton-S	126-75-0	110.0	1234.6	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.23
[313]	Mouse	FMO	8.1	demeton-O	298-03-3	59.3	1771.5	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.77
[313]	Mouse	FMO	8.1	sulprofos	35400-43-2	1.2	728.7	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.73
[313]	Mouse	FMO	8.1	phorate	298-02-2	32.2	1408.0	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.41
[313]	Mouse	FMO	8.1	phorate oxon	2600-69-3	461.7	1170.6	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.17
[313]	Mouse	FMO	8.1	fenthion	55-38-9	12.0	673.3	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.67
[313]	Mouse	FMO	8.1	thiofanox	39196-18-4	574.9	1306.9	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.31
[313]	Mouse	FMO	8.1	methiocarb	2032-65-7	129.9	250.5	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.25
[313]	Mouse	FMO	8.1	aldicarb	116-06-3	607	1087.7	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.09
[313]	Mouse	FMO	8.1	metam-sodium	137-42-8	572	581.1	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.58
[313]	Mouse	FMO	8.1	sodium dimethyl- dithiocarbamate	128-04-1	761.3	1359.8	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.36
[313]	Mouse	FMO	8.1	sodium diethyl- dithiocarbamate	148-18-5	738.9	1099.2	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.10
[313]	Mouse	FMO	8.1	dazomet	533-74-4	398.7	1409.8	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	1.41
[313]	Pig	FMO	8.1	thiourea	62-56-6	49.2	689.1	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.69
[313]	Pig	FMO	8.1	disulfoton	298-04-4	2.2	726.7	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.73



[313]	Pig	FMO	8.1	37	demeton-S	126-75-0	40.4	399.1	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.40
[313]	Pig	FMO	8.1	37	demeton-O	298-03-3	18.2	528.7	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.53
[313]	Pig	FMO	8.1	37	sulprofos	35400-43-2	1.1	607.1	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.61
[313]	Pig	FMO	8.1	37	phorate	298-02-2	12.3	749.4	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.75
[313]	Pig	FMO	8.1	37	phorate oxon	2600-69-3	225.6	640.2	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.64
[313]	Pig	FMO	8.1	37	fenthion	55-38-9	9	464.6	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.46
[313]	Pig	FMO	8.1	37	terbufos	13071-79-9	14.7	285.0	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.29
[313]	Pig	FMO	8.1	37	thiofanox	39196-18-4	136.7	601.4	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.60
[313]	Pig	FMO	8.1	37	methiocarb	2032-65-7	115.2	214.6	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.21
[313]	Pig	FMO	8.1	37	aldicarb	116-06-3	312.5	325.4	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.33
[313]	Pig	FMO	8.1	37	CDEC (sulfallate)	95-06-7	7.4	170.4	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.17
[313]	Pig	FMO	8.1	37	metam-sodium	137-42-8	175.2	266.0	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.27
[313]	Pig	FMO	8.1	37	sodium dimethyl-dithiocarbamate	128-04-1	205.4	500.8	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.50
[313]	Pig	FMO	8.1	37	sodium diethyl-dithiocarbamate	148-18-5	189.3	461.3	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.46
[313]	Pig	FMO	8.1	37	dazomet	533-74-4	244.4	480.0	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.48
[314]	Pig	FMO	8.3	38	chlorpromazine	50-53-3	9.0	2.21	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	2.21
[314]	Pig	FMO	8.3	38	fluphenazine	69-23-8	12.0	2.26	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	2.26
[314]	Pig	FMO	8.3	38	prochlorperazine	58-38-8	3.4	1.70	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.70
[314]	Pig	FMO	8.3	38	thiopropazine	316-81-4	11.0	1.83	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.83
[314]	Pig	FMO	8.3	38	trifluoperazine	117-89-5	13.0	2.54	μmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	2.54
[85]	Pig	FMO	7.5	37	benzenecarbothioic acid	121-68-6	3.3			na
[85]	Pig	FMO	7.5	37	2-hydroxybenzenecarbothioic acid	527-89-9	3.1			na
[85]	Pig	FMO	7.5	37	4-(dimethylamino)benzenecarbo	na	14			na

[85]	Pig	FMO	7.5	37	dithioic acid	na	460	35	na
[85]	Pig	FMO	7.5	37	methyl 2-hydroxybenzenecarbodithioate	na	65	35	na
[85]	Pig	FMO	7.5	37	methyl 4-(dimethylamino)benzenecarbo dithioate	na	890	35	na
[85]	Pig	FMO	7.5	37	3-(dimethylamino)propyl benzenecarbodithioate	na	13	49-51	na
[85]	Pig	FMO	7.5	37	3-(dimethylamino)propyl 2-hydroxybenzenecarbodithioate	na	12	49-51	na
[85]	Pig	FMO	7.5	37	3-(dimethylamino)propyl 4-(dimethylamino)benzenecarbo dithioate	na	23	49-51	na
[85]	Pig	FMO	7.5	37	[(phenylcarbonothioyl)sulfanyl] acetic acid	942-91-6	7500	35	na
[85]	Pig	FMO	7.5	37	{[(2-hydroxyphenyl)carbonothioyl]sulfanyl}acetic acid	na	5400	35	na
[85]	Pig	FMO	7.5	37	2-mercaptobenzoic acid	147-93-3	67	35	na
[85]	Pig	FMO	7.5	37	mercaptoethanol	na	200	35	na
[85]	Pig	FMO	7.5	37	(methylsulfanyl)acetonitrile	35120-10-6	780	35	na
[85]	Pig	FMO	7.5	37	4-(1,2-dithiolan-3-yl)butanoic acid	na	140	35	na
[85]	Pig	FMO	7.5	37	thioctic acid	1077-28-7	120	35	na
[85]	Pig	FMO	7.5	37	thioctamide	940-69-2	2	35	na
[85]	Pig	FMO	7.5	37	6-(1,2-dithiolan-3-yl)hexanoic acid	na	430	35	na
[85]	Pig	FMO	7.5	37	dihydrolipoic acid	462-20-4	1700	35	na
[85]	Pig	FMO	7.5	37	dihydrolipoamide	3884-47-7	15	35	na
[315]	Mouse	FMO	8.4	37	phorate	298-02-2	13.7	11.8 nmol min <sup>-1</sup> mg <sub>m<sup>icr</sup></sub> <sup>-1</sup>	0.41

[315]	Mouse	FMO	8.4	37	disulfoton	298-04-4	3.5	14.5	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	29	0.51
[315]	Mouse	FMO	8.4	37	demeton-S-methyl	919-86-8	34.5	15.9	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	31.8	0.56
[315]	Mouse	FMO	8.4	37	demeton-O	298-03-3	14.8	14.4	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	28.8	0.51
[315]	Mouse	FMO	8.4	37	demeton-S	126-75-0	36.1	13.1	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	26.2	0.46
[315]	Mouse	FMO	8.4	37	sulprofos	35400-43-2	3.2	6.7	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	13.4	0.24
[315]	Mouse	FMO	8.4	37	fenthion	55-38-9	3.4	6.2	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	12.4	0.22
[315]	Mouse	FMO	8.4	37	fosthietan	21548-32-3	80.8	19.4	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	38.8	0.68
[315]	Mouse	FMO	8.4	37	fonofos	944-22-9	13.7	5.6	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	11.2	0.20
[315]	Mouse	FMO	8.4	37	aldicarb	116-06-3	279.0	7.9	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	15.8	0.28
[315]	Mouse	FMO	8.4	37	ethiofencarb	29973-13-5	288.0	15.1	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	30.2	0.53
[315]	Mouse	FMO	8.4	37	nicotine	54-11-5	720.0	10.6	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	21.2	0.37
[315]	Mouse	FMO	8.4	37	ethylenethiourea	96-45-7	253.0	24.2	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	48.4	0.85
[316]	Mouse	FMO	8.4	37	thiourea	62-56-6	20	18.9	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	37.8	0.66
[316]	Mouse	FMO	8.4	37	methimazole	60-56-0	17	17.4	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	34.8	0.61
[316]	Mouse	FMO	8.4	37	cysteamine	60-23-1	105	17.8	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	35.6	0.62
[316]	Mouse	FMO	8.4	37	dimethylaniline	121-69-7	196	16.2	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	32.4	0.57
[316]	Mouse	FMO	8.4	37	trimethylamine	75-50-3	57	16.4	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	32.8	0.58
[316]	Pig	FMO	8.4	37	thiourea	62-56-6	21	12.9	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	25.8	0.46
[316]	Pig	FMO	8.4	37	methimazole	60-56-0	13	14.3	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	28.6	0.51
[316]	Pig	FMO	8.4	37	cysteamine	60-23-1	48	13.6	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	27.2	0.49
[316]	Pig	FMO	8.4	37	dimethylaniline	121-69-7	7	12.2	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	24.4	0.44
[316]	Pig	FMO	8.4	37	trimethylamine	75-50-3	65	10.0	nmol min <sup>-1</sup> mg <sub>micr</sub> <sup>-1</sup>	20	0.36
[317]	Mouse	FMO	8.5	37	thiourea	62-56-6	8.8	1430	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.43	
[317]	Mouse	FMO	8.5	37	methimazole	60-56-0	10.9	1050	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.05	
[317]	Mouse	FMO	8.5	37	thiobenzamide	2227-79-4	28.6	1320	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.32	

[317]	Mouse	FMO	8.5	37	dimethylaniline	121-69-7	116.0	1250	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.25
[317]	Mouse	FMO	8.5	37	imipramine	50-49-7	47.6	1290	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.29
[317]	Mouse	FMO	8.5	37	chlorpromazine	50-53-3	2.3	2330	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	2.33
[317]	Mouse	FMO	8.5	37	nicotine	54-11-5	250.0	410	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.41
[317]	Mouse	FMO	8.5	37	thioridazine	50-52-2	50.0	950	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.95
[317]	Mouse	FMO	8.5	37	diethylphenylphosphine	1605-53-4	2.0	1860	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.86
[317]	Mouse	FMO	8.5	37	fonofos	944-22-9	6.8	270	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.27
[317]	Mouse	FMO	8.5	37	disulfoton	298-04-4	1.4	950	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.95
[317]	Mouse	FMO	8.5	37	fenthion	55-38-9	23.2	1000	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.00
[317]	Mouse	FMO	8.5	37	phorate	298-02-2	18.9	1030	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.03
[317]	Mouse	FMO	8.5	37	aldicarb	116-06-3	196.0	710	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.71
[318]	Pig	FMO	8	322 5	dimethylaniline	121-69-7	18	321	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.32
[318]	Pig	FMO	8	32 25	MPTP (1 -methyl-4-phenyl-1, 2 ,3 ,6-tetrahydropyridine)	28289-54-5	38	495	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.50
[318]	Pig	FMO	8	32 25	1,2,3,4-tetrahydroisoquinoline	91-21-4	1724	270	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.27
[318]	Pig	FMO	8	322 5	1-methyl-6,7- dihydroxytetrahydroisoquinolin e	525-72-4	3571	1064	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	1.06
[319]	Pig	FMO	8	32	dimethylaniline	121-69-7	17	7	nmol min <sup>-1</sup> mg <sub>nicr</sub> <sup>-1</sup>	0.25
[319]	Pig	FMO	8	32	MPTP (1 -methyl-4-phenyl-1, 2 ,3 ,6-tetrahydropyridine)	28289-54-5	47	10	nmol min <sup>-1</sup> mg <sub>nicr</sub> <sup>-1</sup>	0.36
[319]	Pig	FMO	8	32	1,2,3,4-tetrahydroisoquinoline	91-21-4	6900	35	nmol min <sup>-1</sup> mg <sub>nicr</sub> <sup>-1</sup>	1.25
[319]	Pig	FMO	8	32	1-methyl-6,7- dihydroxytetrahydroisoquinolin e	525-72-4	5600	16	nmol min <sup>-1</sup> mg <sub>nicr</sub> <sup>-1</sup>	0.57
[320]	Pig	FMO	8	25	lidocaine	137-58-6	143	145	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.15
[320]	Pig	FMO	8	25	bupivacaine	2180-92-9	408	119	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.12

[320]	Pig	FMO	8	25	propranolol	525-66-6	210	135	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.14
[86]	Pig	FMO			promazine	58-40-2	66			na
[86]	Pig	FMO			triflupromazine	146-54-3	20			na
[86]	Pig	FMO			brompheniramine	86-22-6	200			na
[86]	Pig	FMO			diphenhydramine	58-73-1	160			na
[86]	Pig	FMO			methapyrilene	91-80-5	93			na
[86]	Pig	FMO			benzphetamine	156-08-1	130			na
[86]	Pig	FMO			dimethylaniline	121-69-7	3			na
[86]	Pig	FMO			fluphenazine	69-23-8	12			na
[86]	Pig	FMO			guanethidine	55-65-2	170			na
[86]	Pig	FMO			desipramine	50-47-5	250			na
[86]	Pig	FMO			nortriptyline	72-69-5	500			na
[86]	Pig	FMO			N-methylaniline	100-61-8	30			na
[86]	Pig	FMO			N-methyloctylamine	2439-54-5	400			na
[86]	Pig	FMO			perazine	84-97-9	8000			na

## CYP

Ref.	Species	Isoenz	pH	T	Compound name	CAS	$K_m, \mu\text{M}$	$V_{\text{max}}$	$V_{\text{max}}$ units	$k_{\text{cat}}$ $\text{min}^{-1}$	$V_{\text{max}}$ $\mu\text{mol min}^{-1} \text{mg}_{\text{prot}}^{-1}$
[321]	Rat	CYP1A1	7.4	37	paracetamol	103-90-2	730	0.66	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.00
[321]	Rat	CYP1A1	7.4	37	3,5-dimethyl-4-hydroxyacetanilide	22900-79-4	130	3.00	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.00
[321]	Rat	CYP1A1	7.4	37	3,5-diethyl-4-hydroxyacetanilide	55205-89-5	70	1.70	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.00
[321]	Rat	CYP1A1	7.4	37	3,5-dipropyl-4-hydroxyacetanilide	na	210	1.80	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$		0.00

[321]	Rat	CYP1A1	7.4	37	3,5-difluoro-4-hydroxyacetanilide	na	640	0.82	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.00
[321]	Rat	CYP1A1	7.4	37	3,5-dichloro-4-hydroxyacetanilide	na	160	0.77	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.00
[321]	Rat	CYP1A1	7.4	37	3,5-dibromo-4-hydroxyacetanilide	na	100	0.85	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.00
[321]	Rat	CYP1A1	7.4	37	3,5-diiodo-4-hydroxyacetanilide	na	70	1.70	$\text{nmol min}^{-1} \text{mg}_{\text{prot}}^{-1}$	0.00
[322]	Rabbit	CYP2B4	7		4-xylene	106-42-3	1500		8.17	0.13
[322]	Rabbit	CYP2B4	7		4-fluorotoluene	352-32-9	2200		3.17	0.05
[322]	Rabbit	CYP2B4	7		4-chlorotoluene	106-43-4	580		1.92	0.03
[322]	Rabbit	CYP2B4	7		4-bromotoluene	106-38-7	380		1.67	0.03
[322]	Rabbit	CYP2B4	7		4-tolunitrile	104-85-8	3800		0.37	0.01
[322]	Rabbit	CYP2B4	7		4-nitrotoluene	99-99-0	3300		0.22	0.00
[322]	Rabbit	CYP2B4	7		toluene	108-88-3	4200		4.00	0.06
[322]	Rabbit	CYP2B4	7		3-xylene	108-38-3	1300		4.58	0.07
[322]	Rabbit	CYP2B4	7		3-fluorotoluene	352-70-5	2000		3.17	0.05
[322]	Rabbit	CYP2B4	7		3-chlorotoluene	108-41-8	500		2.92	0.05
[322]	Rabbit	CYP2B4	7		3-bromotoluene	591-17-3	500		2.67	0.04
[322]	Rabbit	CYP2B4	7		3-tolunitrile	620-22-4	5200		2.42	0.04
[322]	Rabbit	CYP2B4	7		3-nitrotoluene	99-08-1	2500		2.08	0.03
[31]	Rat	CYP1A1/1A2	7.6	37	aniline	62-53-3	17000		5.60	0.09
[31]	Rat	CYP1A1/1A2	7.6	37	2-fluoroaniline	348-54-9	7400		6.20	0.10
[31]	Rat	CYP1A1/1A2	7.6	37	2-chloroaniline	95-51-2	600		3.50	0.06
[31]	Rat	CYP1A1/1A2	7.6	37	2-bromoaniline	615-36-1	400		2.80	0.05
[31]	Rat	CYP1A1/1A2	7.6	37	2-iodoaniline	615-43-0	200		2.70	0.05



[323]	Rat	CYP2B1	7.4	37	4-ethyl-N,N-dimethylaniline	4150-37-2	200	89.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.09
[323]	Rat	CYP2B1	7.4	37	4-bromo-N,N-dimethylaniline	586-77-6	81	18.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.02
[323]	Rat	CYP2B1	7.4	37	methyl 4-dimethylaminobenzoate	1202-25-1	76	6.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	4-chloro-N,N-dimethylaniline	698-69-1	235	38.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.04
[323]	Rat	CYP2B1	7.4	37	3-chloro-N,N-dimethylaniline	6848-13-1	1200	6.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	3-iodo-N,N-dimethylaniline	na	30300	7.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	3-nitro-N,N-dimethylaniline	619-31-8	29500	7.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	3-fluoro-N,N-dimethylaniline	na	4300	8.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	3-methoxy-N,N-dimethylaniline	15799-79-8	21300	5.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	3-acetamido-N,N-dimethylaniline	7474-95-5	680	7.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	3-ethyl-N,N-dimethylaniline	na	473	10.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	3-methyl-N,N-dimethylaniline	121-72-2	274	5.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	3-bromo-N,N-dimethylaniline	16518-62-0	51000	6.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	2-chloro-N,N-dimethylaniline	698-01-1	127	7.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.01
[323]	Rat	CYP2B1	7.4	37	4-fluoro-N-methylaniline	459-59-6	133	59.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.06
[323]	Rat	CYP2B1	7.4	37	N-methylaniline	100-61-8	11	19.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.02
[323]	Rat	CYP2B1	7.4	37	4-chloro-N-methylaniline	932-96-7	69	21.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.02
[323]	Rat	CYP2B1	7.4	37	4-methoxy-N-methylaniline	5961-59-1	60	26.00	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>	0.03
[324]	Rat	CYP2B1	7.7	37	3-methyl-N,N-dimethylaniline	121-72-2	150			14.7
[324]	Rat	CYP2B1	7.7	37	4-methyl-N,N-dimethylaniline	99-97-8	160			12.7
[324]	Rat	CYP2B1	7.7	37	N,N-dimethylaniline	121-69-7	260			10.6
[324]	Rat	CYP2B1	7.7	37	4-fluoro-N,N-dimethylaniline	403-46-3	80			9.2
[324]	Rat	CYP2B1	7.7	37	4-chloro-N,N-dimethylaniline	698-69-1	70			9
[324]	Rat	CYP2B1	7.7	37	4-bromo-N,N-dimethylaniline	586-77-6	42			7.5



[324]	Rat	CYP2B1	7.7	37	4-formyl-N,N-dimethylaniline	100-10-7	370				4.3	0.08
[324]	Rat	CYP2B1	7.7	37	4-cyano-N,N-dimethylaniline	1197-19-9	70				2.7	0.05
[324]	Rat	CYP2B1	7.7	37	4-nitro-N,N-dimethylaniline	100-23-2	30				1.4	0.02
[325]	Rat	CYP2B1	7.4	37	N,N-dimethyl-β-naphthylamine	2436-85-3	2.3	172	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.17
[325]	Rat	CYP2B1	7.4	37	3-chloro-N,N-dimethylaniline	6848-13-1	11.5	109	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.11
[325]	Rat	CYP2B1	7.4	37	3-methyl-N,N-dimethylaniline	121-72-2	18.6	220	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.22
[325]	Rat	CYP2B1	7.4	37	4-methyl-N,N-dimethylaniline	99-97-8	20.0	177	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.18
[325]	Rat	CYP2B1	7.4	37	pentobarbital	76-74-4	33.1	185	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.19
[325]	Rat	CYP2B1	7.4	37	hexobarbital	56-29-1	60.3	311	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.31
[325]	Rat	CYP2B1	7.4	37	N,N-dimethylaniline	121-69-7	64.6	195	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.20
[325]	Rat	CYP2B1	7.4	37	codeine	76-57-3	436.5	376	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.38
[325]	Rat	CYP2B1	7.4	37	4-amino-N,N-dimethylaniline	99-98-9	134.9	151	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.15
[325]	Rat	CYP2B1	7.4	37	3-amino-N,N-dimethylaniline	2836-04-6	141.3	136	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.14
[325]	Rat	CYP2B1	7.4	37	ephedrine	299-42-3	10715.2	569	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.57
[325]	Rat	CYP2B1	7.4	37	barbital	57-44-3	1698.2	161	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.16
[325]	Rat	CYP2B1	7.4	37	physostigmine	57-47-6	1148.2	251	nmol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.25
[325]	Rat	CYP2B1	7.4	37	caffeine	58-08-2	1380.4	298	mol min <sup>-1</sup> mg <sub>prot</sub> <sup>-1</sup>			0.30
[326]	Rabbit	CYP2E1	7.6	30	methanol	67-56-1	35000	1.80	mol min <sup>-1</sup> mol <sub>enz</sub> <sup>-1</sup>			0.03
[326]	Rabbit	CYP2E1	7.6	30	ethanol	64-17-5	19000	13.70	mol min <sup>-1</sup> mol <sub>enz</sub> <sup>-1</sup>			0.26
[326]	Rabbit	CYP2E1	7.6	30	1-propanol	71-23-8	7500	15.4	mol min <sup>-1</sup> mol <sub>enz</sub> <sup>-1</sup>			0.29
[326]	Rabbit	CYP2E1	7.6	30	1-butanol	71-36-3	4400	16.40	mol min <sup>-1</sup> mol <sub>enz</sub> <sup>-1</sup>			0.31
[327]	Rabbit	CYP2E1	7.6	30	methyl formate	107-31-3	50000	3.40	mol min <sup>-1</sup> mol <sub>enz</sub> <sup>-1</sup>			0.06
[327]	Rabbit	CYP2E1	7.6	30	methyl acetate	79-20-9	25000	1.40	mol min <sup>-1</sup> mol <sub>enz</sub> <sup>-1</sup>			0.03
[327]	Rabbit	CYP2E1	7.6	30	methyl propionate	554-12-1	20000	0.70	mol min <sup>-1</sup> mol <sub>enz</sub> <sup>-1</sup>			0.01
[327]	Rabbit	CYP2E1	7.6	30	methyl butyrate	623-42-7	4000	0.50	mol min <sup>-1</sup> mol <sub>enz</sub> <sup>-1</sup>			0.01

[327]	Rabbit	CYP2E1	7.6	30	methyl valerate	624-24-8	4000	0.30	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.01
[327]	Rabbit	CYP2E1	7.6	30	ethyl formate	109-94-4	3000	0.80	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.02
[327]	Rabbit	CYP2E1	7.6	30	ethyl acetate	141-78-6	2000	1.00	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.02
[327]	Rabbit	CYP2E1	7.6	30	ethyl propionate	105-37-3	1200	0.60	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.01
[327]	Rabbit	CYP2E1	7.6	30	ethyl butyrate	105-54-4	200	0.30	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.01
[327]	Rabbit	CYP2E1	7.6	30	ethyl valerate	539-82-2	300	0.30	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.01
[327]	Rabbit	CYP2E1	7.6	30	ethyl caproate	123-66-0	100	0.60	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.01
[327]	Rabbit	CYP2E1	7.6	30	ethyl heptanoate	106-30-9	40	0.50	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.01
[327]	Rabbit	CYP2E1	7.6	30	n-propyl acetate	109-60-4	400	0.30	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.01
[327]	Rabbit	CYP2E1	7.6	30	n-butyl acetate	123-86-4	1500	0.15	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.00
[327]	Rabbit	CYP2E1	7.6	30	n-amyl acetate	628-63-7	500	0.05	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.00
[327]	Rabbit	CYP2B4	7.6	30	ethyl formate	109-94-4	35000	3.20	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.05
[327]	Rabbit	CYP2B4	7.6	30	ethyl acetate	141-78-6	37000	14.50	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.23
[327]	Rabbit	CYP2B4	7.6	30	ethyl propionate	105-37-3	24000	9.50	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.15
[327]	Rabbit	CYP2B4	7.6	30	ethyl butyrate	105-54-4	10000	3.20	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.05
[327]	Rabbit	CYP2B4	7.6	30	ethyl valerate	539-82-2	400	4.70	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.07
[327]	Rabbit	CYP2B4	7.6	30	ethyl caproate	123-66-0	300	16.00	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.25
[327]	Rabbit	CYP2B4	7.6	30	ethyl heptanoate	106-30-9	200	10.50	$\text{mol min}^{-1} \text{mol}_{\text{enz}}^{-1}$	0.16
[328]	Rabbit	CYP2E1	7.4	30	4-methoxybenzyl alcohol	105-13-5	1000	4.20		0.08
[328]	Rabbit	CYP2E1	7.4	30	4-methylbenzyl alcohol	589-18-4	310	3.37		0.06
[328]	Rabbit	CYP2E1	7.4	30	benzyl alcohol	100-51-6	450	3.59		0.07
[328]	Rabbit	CYP2E1	7.4	30	4-fluorobenzyl alcohol	459-56-3	320	2.89		0.05
[328]	Rabbit	CYP2E1	7.4	30	4-bromobenzyl alcohol	873-75-6	130	2.08		0.04
[328]	Rabbit	CYP2E1	7.4	30	4-chlorobenzyl alcohol	873-76-7	110	2.37		0.04
[328]	Rabbit	CYP2E1	7.4	30	4-cyanobenzyl alcohol	874-89-5	690	2.31		0.04

[328]	Rabbit	CYP2E1	7.4	30	4-nitrobenzyl alcohol	619-73-8	430	2.90	0.05
[328]	Rabbit	CYP2B4	7.4	30	4-methoxybenzyl alcohol	105-13-5	4310	1.20	0.02
[328]	Rabbit	CYP2B4	7.4	30	4-methylbenzyl alcohol	589-18-4	330	0.79	0.01
[328]	Rabbit	CYP2B4	7.4	30	benzyl alcohol	100-51-6	7280	3.38	0.05
[328]	Rabbit	CYP2B4	7.4	30	4-fluorobenzyl alcohol	459-56-3	1280	0.61	0.01
[328]	Rabbit	CYP2B4	7.4	30	4-bromobenzyl alcohol	873-75-6	130	0.75	0.01
[328]	Rabbit	CYP2B4	7.4	30	4-chlorobenzyl alcohol	873-76-7	130	0.52	0.01
[328]	Rabbit	CYP2B4	7.4	30	4-cyanobenzyl alcohol	874-89-5	1480	0.49	0.01
[328]	Rabbit	CYP2B4	7.4	30	4-nitrobenzyl alcohol	619-73-8	1900	0.73	0.01
[328]	Rabbit	CYP2E1	7.4	30	1-(4-methoxyphenyl)ethanol	3319-15-1	640	3.50	0.07
[328]	Rabbit	CYP2E1	7.4	30	1-(4-methylphenyl)ethanol	536-50-5	440	3.38	0.06
[328]	Rabbit	CYP2E1	7.4	30	1-phenylethanol	98-85-1	490	4.04	0.08
[328]	Rabbit	CYP2E1	7.4	30	1-(4-fluorophenyl)ethanol	403-41-8	260	2.35	0.04
[328]	Rabbit	CYP2E1	7.4	30	1-(4-bromophenyl)ethanol	5391-88-8	60	0.92	0.02
[328]	Rabbit	CYP2E1	7.4	30	1-(4-chlorophenyl)ethanol	3391-10-4	50	1.76	0.03
[328]	Rabbit	CYP2E1	7.4	30	4-(1-hydroxyethyl)benzoic acid	na	7540	0.62	0.01
[328]	Rabbit	CYP2E1	7.4	30	4-(1-hydroxyethyl)benzonitrile	na	420	1.19	0.02
[328]	Rabbit	CYP2E1	7.4	30	1-(4-nitrophenyl)ethanol	na	130	1.06	0.02
[328]	Rabbit	CYP2B4	7.4	30	1-(4-methoxyphenyl)ethanol	3319-15-1	3120	3.46	0.05
[328]	Rabbit	CYP2B4	7.4	30	1-(4-methylphenyl)ethanol	536-50-5	720	4.37	0.07
[328]	Rabbit	CYP2B4	7.4	30	1-phenylethanol	98-85-1	1880	6.03	0.09
[328]	Rabbit	CYP2B4	7.4	30	1-(4-fluorophenyl)ethanol	403-41-8	650	2.28	0.04
[328]	Rabbit	CYP2B4	7.4	30	1-(4-bromophenyl)ethanol	5391-88-8	60	1.59	0.02
[328]	Rabbit	CYP2B4	7.4	30	1-(4-chlorophenyl)ethanol	3391-10-4	270	1.93	0.03
[328]	Rabbit	CYP2B4	7.4	30	4-(1-hydroxyethyl)benzoic acid	na	6090	0.31	0.00

[328]	Rabbit	CYP2B4	7.4	30	4-(1-hydroxyethyl)benzonitrile	na	1190		0.97	0.02
[328]	Rabbit	CYP2B4	7.4	30	1-(4-nitrophenyl)ethanol	na	2310		1.00	0.02
[329]	Rabbit	CYP2B4	7.7	T <sub>room</sub>	4-methoxy thioanisole	1879-16-9	63	18 mol min <sup>-1</sup> mol <sub>CYP</sub> <sup>-1</sup>		0.29
[329]	Rabbit	CYP2B4	7.7	T <sub>room</sub>	4-methylthioanisole	623-13-2	77	17 mol min <sup>-1</sup> mol <sub>CYP</sub> <sup>-1</sup>		0.27
[329]	Rabbit	CYP2B4	7.7	T <sub>room</sub>	thioanisole	100-68-5	110	14 mol min <sup>-1</sup> mol <sub>CYP</sub> <sup>-1</sup>		0.22
[329]	Rabbit	CYP2B4	7.7	T <sub>room</sub>	4-chlorothioanisole	123-09-1	135	25 mol min <sup>-1</sup> mol <sub>CYP</sub> <sup>-1</sup>		0.40
[329]	Rabbit	CYP2B4	7.7	T <sub>room</sub>	4-nitrothioanisole	701-57-5	31	11 mol min <sup>-1</sup> mol <sub>CYP</sub> <sup>-1</sup>		0.17
[330]	Rabbit	CYP2B4	7.4	24.5	4-methoxyphenyl methyl sulfoxide	na	78	3.2 mol min <sup>-1</sup> mol <sub>CYP</sub> <sup>-1</sup>		0.05
[330]	Rabbit	CYP2B4	7.4	24.5	methyl 4-methylphenyl sulfoxide	934-72-5	364	2.7 mol min <sup>-1</sup> mol <sub>CYP</sub> <sup>-1</sup>		0.04
[330]	Rabbit	CYP2B4	7.4	24.5	methylphenyl sulfoxide	1193-82-4	74	2.3 mol min <sup>-1</sup> mol <sub>CYP</sub> <sup>-1</sup>		0.04
[330]	Rabbit	CYP2B4	7.4	24.5	4-chlorophenyl methyl sulfoxide	934-73-6	58	2.1 mol min <sup>-1</sup> mol <sub>CYP</sub> <sup>-1</sup>		0.03
[331]	Rabbit	CYP2B4	7.4	25	4-iodotoluene	624-31-7	1050		18.2	0.29
[331]	Rabbit	CYP2B4	7.4	25	4-xylene	106-42-3	1630		13.0	0.21
[331]	Rabbit	CYP2B4	7.4	25	4-bromotoluene	106-38-7	730		11.5	0.18
[331]	Rabbit	CYP2B4	7.4	25	toluene	108-88-3	22000		7.50	0.12
[331]	Rabbit	CYP2B4	7.4	25	4-chlorotoluene	106-43-4	5200		6.10	0.10
[331]	Rabbit	CYP2B4	7.4	25	4-fluorotoluene	352-32-9	1500		4.10	0.07
[331]	Rabbit	CYP2B4	7.4	25	4-tolunitrile	104-85-8	64		2.07	0.03
[331]	Rabbit	CYP2B4	7.4	25	4-nitrotoluene	99-99-0	7.9		1.03	0.02

**Table B2.** Data conversion for rates.

Catalytic rates were reported in the papers with heterogeneous units and with different constants (i.e., as  $V_{\max}$  or as  $k_{\text{cat}}$ ). Therefore, it was necessary to standardise the data. We expressed all rates as  $V_{\max}$ , using  $\mu\text{mol min}^{-1} \text{mg}_{\text{PROT}}^{-1}$  as units. For CYP enzymes,  $V_{\max}$  was referred to the microsomal protein weight, whereas for the other enzymes  $V_{\max}$  was referred to the enzyme weight, i.e.,  $\text{mg}_{\text{PROT}} = \text{mg}_{\text{MICR PROT}}$  and  $\text{mg}_{\text{PROT}} = \text{mg}_{\text{ENZ}}$ , respectively. The rates expressed as  $k_{\text{cat}}$  ( $\text{min}^{-1}$ ) were transformed into  $V_{\max}$  values.

For ADH, ALDH and FMO,  $V_{\max}$  (expressed as  $\mu\text{mol min}^{-1} \text{mg}_{\text{ENZ}}^{-1}$ ) was derived using the molecular weight of the enzyme ( $M_r$ ,  $\mu\text{g}_{\text{ENZ}} \mu\text{mol}^{-1}$ ):

$$V_{\max} = \frac{k_{\text{cat}}}{M_r \cdot 10^{-3}} \quad (\text{eq. 1})$$

$$[\mu\text{mol min}^{-1} \text{mg}_{\text{ENZ}}^{-1}] = \frac{[\text{min}^{-1}]}{[\mu\text{g}_{\text{ENZ}} \mu\text{mol}^{-1}][\text{mg} \mu\text{g}^{-1}]}$$

For CYP, we transformed the  $k_{\text{cat}}$  in  $V_{\max}$  values (expressed as  $\mu\text{mol min}^{-1} \text{mg}_{\text{PROT}}^{-1}$ ) using the specific content of the enzyme ( $E$ ,  $\text{nmol mg}_{\text{MICR PROT}}^{-1}$ ) [29]:

$$\begin{aligned} V_{\max} &= k_{\text{cat}} \cdot [E] \cdot 10^{-3} \quad (\text{eq.2}) \\ &= [\mu\text{mol min}^{-1} \text{mg}_{\text{MICR PROT}}^{-1}] \\ &= [\text{min}^{-1}] \cdot [\text{nmol mg}_{\text{MICR PROT}}^{-1}] \cdot [\mu\text{mol nmol}^{-1}] \end{aligned}$$

In case  $M_r$  or  $[E]$  values were not reported in the paper where we collected  $k_{\text{cat}}$ , we used average values coming from other studies.

The operations performed to standardise the rates are reported in the following table.

## ADH

S.	Species	pH	T	Isoenz.	M <sub>r</sub> enzyme [g <sub>ENZ</sub> <sup>-1</sup> mol <sub>ENZ</sub> <sup>-1</sup> ]	Enzyme abundance [mg <sub>ENZ</sub> g <sup>-1</sup> LIV]	Enzyme conc. [units]	k <sub>cat</sub> or V <sub>max</sub>	Units	Data treatment	# Comp.	Notes
[280]	Human	7.5	25	ADH2	78000 (experim and aa)	Not reported	3 ml assay	V <sub>max</sub>	μmol min <sup>-1</sup> mg <sup>-1</sup> ENZ	/	2	
[281]	Human	7.5 10	25	ADH1	80000 (exp)	Not reported	Not reported	k <sub>cat</sub>	μmol min <sup>-1</sup> μmol <sup>-1</sup> ACT. SITE	→ Multiply by 0.025 <sup>6</sup>	6	
[282]	Human	7.5	25	ADH1	80000 (aa)	0.07 <sup>7</sup>	Not reported	V <sub>max</sub>	μmol min <sup>-1</sup> mg <sup>-1</sup> ENZ	/	27	
[68]	Human	7	25	ADH1, ADH2, ADH3	80000 (exp)	Not reported	Not reported	k <sub>cat</sub>	min <sup>-1</sup>	→ Eq. 1	15	
[270]	Human	10	25	ADH2	80000 (exp)	0.405 <sup>8</sup>	Not reported	k <sub>cat</sub>	min <sup>-1</sup>	→ Eq. 1	13	
[283]	Human Horse	10	25	ADH1, ADH3 ADH1	Not reported	Not reported	Not reported	k <sub>cat</sub>	min <sup>-1</sup>	→ Eq. 1, 80000 as M <sub>r</sub>	31	

<sup>6</sup> The active enzyme has a molecular weight of 80,000 and is a dimer of two identical subunits. Each subunit has one main coenzyme-binding site (Brändén et al., 1973). It means that there are 2 active sites in ADH. [Brändén C-I, Eklund H, Nordström B, Boiwe T, Söderlund G, Zeppezauer E, Ohlsson I, Åkeson Å. 1973. Structure of liver alcohol dehydrogenase at 2.9-Å resolution. Proceedings of the National Academy of Sciences 70:2439-2442.]

In order to express V<sub>max</sub> in [μmol min<sup>-1</sup> mg<sub>ENZ</sub><sup>-1</sup>], the data need to be multiplied by  $2 \frac{\mu\text{mol}_{\text{ACT.SITE}}}{\mu\text{mol}_{\text{ENZYME}}}$  and then divided by M<sub>r</sub> enzyme  $[80,000 \cdot \frac{\text{mg}_{\text{ENZYME}}}{10^3 \mu\text{mol}_{\text{ENZYME}}}] \rightarrow$  multiply the original data by 0.025.

<sup>7</sup> The purification yields 7 mg of homogeneous alcohol dehydrogenase, using 100 g human liver. N.B. 2 chromatography procedures at pH 8.6 and 7.7.

<sup>8</sup> The purification yields 30.375 mg (0.75\*40.5) of homogeneous alcohol dehydrogenase, using 75 g human liver. N.B. chromatography procedures at pH 7.3.

[260]	Horse	7.3	37	ADH	Not reported	Not reported	Not reported	$V_{\max}$	$\text{nmol h}^{-1}\text{mg}^{-1}_{\text{ENZ}}$	Change units	3	Not inserted as results not in line (values close to zero).
[284]	Human Rat	10.5	25	ADH1	Not reported	Not reported	Not reported	$V_{\max}$	$\text{nmol min}^{-1}\text{mg}^{-1}_{\text{ENZ}}$	Change units	2	Not inserted as referred to whole omogenate.
[285]	Rat	10 7.5	25	ADH3 ADH1	80000 exp	0.017	Not reported	$k_{\text{cat}}$	$\text{min}^{-1}$	→ Eq. 1	22	
[286]	Human	6.8 10	25	ADH1	78/79000 (exp) 83000 (aa)	0.32 <sup>9</sup>	3 ml assay	$k_{\text{cat}}$	$\text{min}^{-1}$	→ Eq. 1, 80000 as $M_r$	12	
[287]	Horse	7	25	ADH1	80000 exp	0.32 <sup>10</sup>	Not reported	$k_{\text{cat}}$	$\mu\text{mol s}^{-1}$ $\mu\text{mol}^{-1}_{\text{ACT. SITE}}$	→ Change units and multiply by $0.025^{11}$	13	
[288]	Human	10	25	ADH1	80000 exp	Not reported	Not reported	$k_{\text{cat}}$	$\text{min}^{-1}$	→ Eq. 1	48	
[289]	Human	10 6.8	25	ADH3	82700 exp	0.408 <sup>12</sup>	Not reported	$k_{\text{cat}}$	$\text{min}^{-1}$	→ Eq. 1, 80000 as $M_r$	5	

<sup>9</sup> The purification yields 195 mg of homogeneous alcohol dehydrogenase, using 606 g human liver. N.B. chromatography at pH 7.5

<sup>10</sup> The purification yields 500 mg of homogeneous alcohol dehydrogenase, using 1000 g human liver.

<sup>11</sup> The active enzyme has a molecular weight of 80,000 and is a dimer of two identical subunits, each one having 1 active site. In order to express  $V_{\max}$  in  $[\mu\text{mol min}^{-1}\text{mg}_{\text{ENZ}}^{-1}]$ , the data need to be multiplied by  $2 \frac{\mu\text{mol}_{\text{ACT. SITE}}}{\mu\text{mol}_{\text{ENZYME}}}$  and then divided by  $M_r$ , enzyme  $[80,000 \cdot \frac{\text{mg}_{\text{ENZYME}}}{10^3 \mu\text{mol}_{\text{ENZYME}}}] \rightarrow$  multiply the original data by 0.025.

<sup>12</sup> The purification yields 8.162 mg (0.077\*106) of homogeneous alcohol dehydrogenase, using 200 g human liver.

## ALDH

S.	Species	pH	T	Isoenz.	M <sub>r</sub> enzyme [g <sub>ENZ</sub> mol <sup>-1</sup> <sub>ENZ</sub> ]	Enzyme abundance [mg <sub>ENZ</sub> g <sup>-1</sup> <sub>LIV</sub> ]	Enzyme conc. [units]	k <sub>cat</sub> or V <sub>max</sub>	Units	Data treatment	# Comp.	Notes
[290]	Human	7.0 7.1	25	ALDH1 ALDH2 ALDH3	Not reported	Not reported	Not reported	V <sub>max</sub>	μmol min <sup>-1</sup> mg <sup>-1</sup> <sub>ENZ</sub>	/	30	
[291]	Human	7.4	25	ALDH3	220000	0.0027 <sup>13</sup>	Not reported	V <sub>max</sub>	μmol min <sup>-1</sup> mg <sup>-1</sup> <sub>ENZ</sub>	/	1	
[292]	Horse	7.0	25	ALDH1	230000	0.25	Not reported	V	V relative to acetaldehyde	Multiply by 0.13	19	
				ALDH2	240000	0.125 <sup>14</sup>			Multiply by 0.35 <sup>15</sup>			
[293]	Rat Human	7.4 9.0	25	ALDH2	220000	Not reported	Not reported	k <sub>cat</sub>	min <sup>-1</sup>	→ Eq. 1	3	
[294]	Human	9.5	(25)	ALDH1	245000	0.03	Not reported	V <sub>max</sub>	μmol min <sup>-1</sup> mg <sup>-1</sup> <sub>ENZ</sub>	/	4	
				ALDH2	225000	0.08 <sup>16</sup>						
[295]	Rat	8.0	25	ALDH1	237000	2.95	9 [mg <sub>ENZ</sub> ml <sup>-1</sup> ]	V <sub>max</sub>	μmol min <sup>-1</sup> ml <sup>-1</sup>	→ Divide by enzyme conc.	8	
				ALDH1	234000	2.59	Not reported	V <sub>max</sub>	μmol min <sup>-1</sup> mg <sup>-1</sup> <sub>ENZ</sub>	/	7	
				ALDH2	204000	3.89 <sup>17</sup>						

<sup>13</sup> The purification yields 0.27 mg of homogeneous ALDH3 using 100 g human liver.

<sup>14</sup> The purification yields 200 mg of homogeneous ALDH1 and 100 mg of ALDH2, using 800 g horse liver.

<sup>15</sup> These values (0.13 and 0.35) are the specific activities of ALDH1 and ALDH2 [ $\mu\text{mol min}^{-1}\text{mg}^{-1}_{\text{ENZ}}$ ] with 1mM acetaldehyde.

<sup>16</sup> The purification yields 20 mg of ALDH1 and 50 mg of ALDH2, using 600 g human liver.

<sup>17</sup> The purification yields 9 mg of ALDH1 using 3.047 g rat liver, and 12 mg of ALDH1 and 18mg of ALDH2 using 4.628 g rat liver.



[296]	Human	7.4 and 9	25	ALDH1	Not reported	3	$V_{\max}$	$\mu\text{mol min}^{-1} \text{mg}^{-1} \text{ENZ}$	/	15	
			25	ALDH2	Not reported						
				ALDH3	216000 <sup>18</sup>						
[80]	Human	7.4	25	ALDH1	Not reported	Not reported	/			5	$V_{\max}$ or $k_{\text{cat}}$ not reported
[70]	Human	9.5	25	ALDH1	230000	Not reported	$k_{\text{cat}}$	$\text{min}^{-1}$	$\rightarrow$ Eq. 1	67	
				ALDH2	240000						
[297]	Human	7.4	25	ALDH3	219000	0.0105 <sup>19</sup>	$V_{\max}$	$\mu\text{mol min}^{-1} \text{mg}^{-1} \text{ENZ}$	/	6	
[298]	Rat	8.5	25	ALDH1	170000	Not reported	V	V relative to propionaldehyde	Multiply by 4.348 or 4.599	8	
				ALDH2	250000				Multiply by 9.928 or 5.689 <sup>20</sup>		
[299]	Rat	7.4	37	ALDH1	216000 <sup>21</sup>	Not reported	$V_{\max}$	$\text{nmol min}^{-1} \text{mg}^{-1} \text{ENZ}$	Change units	1	
[300]	Human	7.5 and 9.5	25	ALDH1	245000	Not reported	$k_{\text{cat}}$	$\text{min}^{-1}$	$\rightarrow$ Eq. 1	4	
				ALDH2	225000						

<sup>18</sup> Apparent subunit molecular mass of 54 kDa. ALDH is a tetrameric enzyme  $\rightarrow$  ALDH  $M_r = 54000 \times 4$ .

<sup>19</sup> The purification yields 6.3 mg of homogeneous ALDH3, using 600 g human liver.

<sup>20</sup> These values are the specific activities of ALDH1 and ALDH 2 [ $\mu\text{mol min}^{-1} \text{mg}_{\text{ENZ}}^{-1}$ ] with propionaldehyde [note: 4.348 and 9.928 refer to mitochondria, and 4.599 and 5.689 to microsomes].

<sup>21</sup> Apparent subunit molecular mass of 54 kDa. ALDH is a tetrameric enzyme  $\rightarrow$  ALDH  $M_r = 54000 \times 4$ .

<sup>22</sup> The purification yields 2.4 mg of homogeneous ALDH1 and 8.5 mg of ALDH2, using 50 g human liver.

Continuation of ALDH

[301]	Rat	9.6	37	ALDH1 ALDH2 ALDH3	Not reported	Not reported	3ml assay	$V_{\max}$	$\text{nmol min}^{-1} \text{mg}^{-1} \text{ENZ}$	Change units	23
[302]	Rat	7.4	25	ALDH1	Not reported	Not reported	Not reported	$V_{\max}$	$\text{nmol min}^{-1} \text{mg}^{-1} \text{ENZ}$	Change units	23
[303]	Rat	7.4 (9.6)	25	ALDH1	22.7	8.7 $\text{mg}_{\text{PROT}}$ in 808.2 $\text{mg}_{\text{MITOCOND}}$	Not reported	$V$	$V$ relative to propionaldehyde	Multiply by 0.2317	29
				ALDH2	67000		Not reported			Multiply by 0.7983 <sup>24</sup>	

FMO

S.	Species	pH	T	Isoenz.	$M_r$ enzyme [ $\text{g}_{\text{ENZ}} \text{mol}_{\text{ENZ}}^{-1}$ ]	Enzyme abundance [ $\text{mg}_{\text{ENZ}} \text{g}^{-1} \text{LIV}$ ]	Enzyme conc. [units]	$k_{\text{cat}}$ or $V_{\max}$	Units	Data treatment	# Comp.	Notes
[304]	Pig	7.4	37	FMO <sub>pig</sub>	Not reported	Not reported	Not reported	$V_{\max}$	$\text{nmol min}^{-1} \text{mg}^{-1} \text{ENZ}$	Change units	6	
[305]	Pig	7.5	37	FMO <sub>pig</sub>	Not reported	Not reported	88 [ $\text{pmol}_{\text{ENZ}} \text{ml}^{-1}$ ]	$k_{\text{cat}}$	$\text{min}^{-1}$	→ Eq. 1, <sup>25</sup> 56000 as $M_r$	4	

<sup>23</sup> The  $M_r$  enzyme has been derived from the ratio  $k_{\text{cat}}/V_{\max}$

$$\rightarrow M_r \left[ \frac{\text{g}_{\text{ENZ}}}{\text{mol}} \right] = \frac{k_{\text{cat}} [\text{min}^{-1}]}{V_{\max} [\mu\text{mol} \cdot \text{min}^{-1} \text{mg}_{\text{ENZ}}^{-1}]} \cdot \frac{10^6 [\mu\text{mol} \cdot \text{mol}^{-1}]}{10^3 [\text{mg} \cdot \text{g}^{-1}]} = \frac{123 [\text{min}^{-1}]}{0.56 [\mu\text{mol} \cdot \text{min}^{-1} \text{mg}_{\text{ENZ}}^{-1}]} \cdot \frac{10^6 [\mu\text{mol} \cdot \text{mol}^{-1}]}{10^3 [\text{mg} \cdot \text{g}^{-1}]} = 220000 \left[ \frac{\text{g}_{\text{ENZ}}}{\text{mol}} \right]$$

<sup>24</sup> These values are the specific activities of ALDH1 and ALDH 2 [ $\mu\text{mol min}^{-1} \text{mg}_{\text{ENZ}}^{-1}$ ] with propionaldehyde.

<sup>25</sup> The molecular weight of FMO was taken from Sabourin PJ, Smyser BP, Hodgson E (1984) Int J Biochem 16, 713-720. It is 56000 [ $\text{g}_{\text{ENZ}} \text{mol}^{-1}$ ] for PLG and 57000 [ $\text{g}_{\text{ENZ}} \text{mol}^{-1}$ ] for MOUSE.

[306]	Pig	8.4	37	FMO <sub>pig</sub>	Not reported	Not reported	62.5 [ $\mu\text{g}_{\text{ENZ}}^{-1}\text{ml}^{-1}$ ] <sup>26</sup>	V <sub>max</sub>	nmol min <sup>-1</sup> $\mu\text{g}_{\text{ENZ}}^{-1}$	/	1	
[81]	Pig	7.5	37	FMO <sub>pig</sub>	Not reported	Not reported	Not reported	k <sub>cat</sub>	nmol min <sup>-1</sup> $\text{nmol}_{\text{ENZ}}^{-1}$	→ Eq. 1, 56000 as M <sub>r</sub>	12*	
[307]	Pig	7.4	37	FMO <sub>pig</sub>	Not reported	Not reported	Not reported	V <sub>max</sub>	nmol min <sup>-1</sup> $\text{mg}_{\text{ENZ}}^{-1}$	Change units	3	
[308]	Pig	8.3	37	FMO <sub>pig</sub>	Not reported	Not reported	Not reported	V <sub>max</sub>	$\mu\text{mol min}^{-1}$ $\text{mg}_{\text{ENZ}}^{-1}$	/	6	
[82]	Pig	7.4 and 8.4	37	FMO <sub>pig</sub>	Not reported	Not reported	Not reported	/			28	V <sub>max</sub> or k <sub>cat</sub> not reported
[309]	Pig	8.4	37	FMO <sub>pig</sub>	Not reported	Not reported	0.4 [ $\text{mg}_{\text{ENZ}}^{-1}\text{ml}^{-1}$ ] <sup>27</sup>	V <sub>max</sub>	nmol min <sup>-1</sup> $\text{mg}_{\text{ENZ}}^{-1}$	Change units	6	
[83]	Pig	7.4	37	FMO <sub>pig</sub>	Not reported	Not reported	Not reported	V <sub>max</sub>	0.54-0.56 $\mu\text{mol min}^{-1}$ $\text{mg}_{\text{ENZ}}^{-1}$		3	Exact V <sub>max</sub> not reported
[310]	Pig	7.7	25	FMO <sub>pig</sub>	Not reported	Not reported	0.054 [ $\text{mg}_{\text{ENZ}}^{-1}\text{ml}^{-1}$ ]	V <sub>max</sub>	nmol min <sup>-1</sup> $\text{mg}_{\text{ENZ}}^{-1}$	Change units	5	
[311]	Pig	8.2	37	FMO <sub>pig</sub>	Not reported	Not reported	Not reported	V <sub>max</sub>	nmol min <sup>-1</sup> $\text{mg}_{\text{ENZ}}^{-1}$	Change units	18	
[312]	Mouse	8.1	37	FMO <sub>pig</sub>	Not reported	Not reported	8.8	V <sub>max</sub>	$\mu\text{mol min}^{-1}$ $\text{mg}_{\text{ENZ}}^{-1}$	/	44	
	Pig			FMO <sub>m</sub>			1.7 [ $\mu\text{g}_{\text{ENZ}}^{-1}\text{ml}^{-1}$ ]					

<sup>26</sup> The assay contained 25  $\mu\text{g}$  of purified enzyme in 0.4 ml.

\* k<sub>cat</sub> available for 6 out of 12 compounds.

<sup>27</sup> The assay contained 12 mg of purified enzyme in 30 ml.

Continuation of FMO

[84]	Pig	7.6	37	FMO <sub>pig</sub>	Not reported	Not reported	Not reported	/	/		4	V <sub>max</sub> or k <sub>cat</sub> not reported
[313]	Mouse	8.1	37	FMO <sub>pig</sub>	Not reported	Not reported	7.9 1.8 [μg <sub>ENZ</sub> ml <sup>-1</sup> ]	V <sub>max</sub>	nmol min <sup>-1</sup> mg <sup>-1</sup> ENZ	Change units	32	
	Pig			FMO <sub>mouse</sub>								
[314]	Pig	8.3	38	FMO <sub>pig</sub>	Not reported	Not reported	0.05-0.07 [mg <sub>ENZ</sub> ml <sup>-1</sup> ]	V <sub>max</sub>	μmol min <sup>-1</sup> mg <sup>-1</sup> ENZ	/	5	
[85]	Pig	7.5	37	FMO <sub>pig</sub>	Not reported	Not reported	Not reported	k <sub>cat</sub>	35-51 min <sup>-1</sup>		20	Exact k <sub>cat</sub> not reported (35-51)
[315]	Mouse	8.4	37	FMO <sub>mouse</sub>	Not reported	Not reported	0.1-0.4 [mg <sub>ENZ</sub> ml <sup>-1</sup> ]	V <sub>max</sub>	nmol min <sup>-1</sup> mg <sup>-1</sup> MICR PROT	→ Divide by specific content <sup>28</sup> to obtain k <sub>cat</sub> ; → Eq. 1, <sup>29</sup> 57000 as M <sub>r</sub>	13	

<sup>28</sup> The specific content (0.5 nmol<sub>FMO</sub> mg<sub>MICR PROT</sub><sup>-1</sup>) is the value measured for pig in Dannan GA and Guengerich FP (1982) Mol Pharmacol 22, 787-794. No value for pig was found.

<sup>29</sup> The molecular weight of FMO was taken from Sabourin PJ, Smyser BP, Hodgson E (1984) Int J Biochem 16, 713-720. It is 56000 [g<sub>ENZ</sub> mol<sup>-1</sup>] for PIG and 57000 [g<sub>ENZ</sub> mol<sup>-1</sup>] for MOUSE.

[316]	Mouse and Pig	8.4	37	FMO <sub>mouse</sub> FMO <sub>pig</sub>	Not reported	Not reported	0.1-0.4 [mg <sub>ENZ</sub> ml <sup>-1</sup> ]	V <sub>max</sub>	nmol min <sup>-1</sup> mg <sup>-1</sup> MICR PROT	→ Divide by specific content <sup>30</sup> to obtain k <sub>cat</sub> ; → Eq. 1, 57 or 56000 as M <sub>r</sub> <sup>31</sup>	10	
[317]	Mouse	8.5	37	FMO <sub>mouse</sub>	58000	0.344 <sup>32</sup>	2.2 [mg <sub>ENZ</sub> ml <sup>-1</sup> ]	V <sub>max</sub>	nmol min <sup>-1</sup> mg <sup>-1</sup> ENZ	Change units	14	
[318]	Pig	8	32 or 25	FMO <sub>pig</sub>	56000	3.997 <sup>33</sup>	Assay 1 or 2.5 ml	V <sub>max</sub>	nmol min <sup>-1</sup> mg <sup>-1</sup> ENZ	Change units	4	
[319]	Pig	8	32	FMO <sub>pig</sub>	Not reported	Not reported	184 [μg <sub>ENZ</sub> ml <sup>-1</sup> ]	V <sub>max</sub>	nmol min <sup>-1</sup> mg <sup>-1</sup> MICR PROT	→ Divide by specific content to obtain k <sub>cat</sub> ; → Eq. 1, 56000 as M <sub>r</sub>	4	
[320]	Pig	8	25	FMO <sub>pig</sub>	Not reported	Not reported	10 [μg <sub>ENZ</sub> ml <sup>-1</sup> ]	V <sub>max</sub>	nmol min <sup>-1</sup> mg <sup>-1</sup> ENZ	Change units	3	
[86]	Pig			FMO <sub>pig</sub>	Not reported	Not reported	Not reported	/			14	V <sub>max</sub> or k <sub>cat</sub> not reported

<sup>30</sup> The specific content (0.5 nmol<sub>FMO</sub> mg<sub>MICR PROT</sub><sup>-1</sup>) is the value measured for pig in Dannan GA and Guengerich FP (1982) Mol Pharmacol 22, 787-794. No value for pig was found.

<sup>31</sup> The molecular weight of FMO was taken from Sabourin PJ, Smyser BP, Hodgson E (1984) Int J Biochem 16, 713-720. It is 56000 [g<sub>ENZ</sub> mol<sup>-1</sup>] for PIG and 57000 [g<sub>ENZ</sub> mol<sup>-1</sup>] for MOUSE.

<sup>32</sup> The purification yields 2.2 mg of homogeneous FMO from 6.4 g of mouse microsomal proteins (200 livers).

<sup>33</sup> The purification yields 19.8 mg of homogeneous FMO from 4.954 g of pig microsomal proteins.

## CYP

S.	Species	pH	T	Isoenz.	E* [nmol <sub>CYP</sub> mg <sup>-1</sup> mg <sup>-1</sup> PROT]	Enzyme conc. [units] [0.1 [mg <sub>MICR</sub> ml <sup>-1</sup> ]	k <sub>cat</sub> or V <sub>max</sub>	Units	Data treatment	# Comp.
[321]	Rat	7.4	37	CYP1A1	Not reported	0.1 [mg <sub>MICR</sub> ml <sup>-1</sup> ]	V <sub>max</sub>	nmol min <sup>-1</sup> mg <sup>-1</sup> PROT	Change units	8
[322]	Rabbit	7	10	CYP2B4	15.4-18.8	Not reported	k <sub>cat</sub>	min <sup>-1</sup>	→ Eq. 2, 15.9 as [E] <sup>34</sup>	13
[31]	Rat	7.6	37	CYP1A1	Not reported	Not reported	k <sub>cat</sub>	min <sup>-1</sup>	→ Eq. 2, 16.9 <sup>35</sup> as [E]	15
[323]	Rat	7.4	37	CYP2B1	Not reported	0.5-2 [mg <sub>MICR</sub> ml <sup>-1</sup> ]	V <sub>max</sub>	nmol min <sup>-1</sup> mg <sup>-1</sup> PROT	Change units	26
[324]	Rat	7.7	37	CYP2B1	Not reported	2 <sup>36</sup> [nmol <sub>CYP</sub> ml <sup>-1</sup> ]	k <sub>cat</sub>	min <sup>-1</sup>	→ Eq. 2, 17.6 <sup>37</sup> as [E]	9
[325]	Rat	7.4	37	CYP2B1	Not reported	0.5 [mg <sub>MICR</sub> ]	V <sub>max</sub>	nmol min <sup>-1</sup> mg <sup>-1</sup> PROT	Change units	14

\* Specific content of the enzyme.

<sup>34</sup> This value (15.9±0.4) is the average value from sources [59-61] (16.4, 15.6 and 15.7, see values in bold on next page). It is in the interval 15.4-18.8 reported in the paper and in accordance with the theoretical specific content (17.7) [Black SD and Coon MJ, Comparative structures of P-450 cytochromes. In *Cytochrome P-450: Structure, Mechanism, and Biochemistry*, Ortiz de Montellano, PR, Ed. Plenum Press: New York, 1986; pp 161-216.].

<sup>35</sup> The value 16.9 is the theoretical specific content from the book by Black SD and Coon MJ (1986). This value was taken as no experimental values were provided in these papers and because the assays were conducted with highly purified enzymes.

<sup>36</sup> The assay contained 0.3 nmol<sub>CYP</sub> in 0.125ml+25µl of solution.

<sup>37</sup> The value 17.6 is the theoretical specific content from the book by Black SD and Coon MJ (1986). This value was taken as no experimental values were provided in these papers and because the assays were conducted with highly purified enzymes.

[326] <sup>38</sup>	Rabbit	7.6	30	CYP2E1	16-20	2 [mg <sub>MICR</sub> ml <sup>-1</sup> ]	k <sub>cat</sub>	nmol min <sup>-1</sup> nmol <sup>-1</sup> <sub>CYP2E1</sub>	→ Eq. 2, 18.8 as [E] <sup>39</sup>	4
[327]	Rabbit	7.6	30	CYP2E1	18.8	0.1 [nmol <sub>CYP</sub> ml <sup>-1</sup> ]	k <sub>cat</sub>	nmol min <sup>-1</sup> nmol <sup>-1</sup> <sub>CYP</sub>	→ Eq. 2, 18.8 as [E]	22
				CYP2B4	<b>15.6</b>				→ Eq. 2, 15.6 as [E]	
[328]	Rabbit	7.4	30	CYP2E1	18.8	0.1/0.5=0.2	k <sub>cat</sub>	min <sup>-1</sup>	→ Eq. 2, 18.8 as [E]	34
				CYP2B4	<b>15.7</b>	0.2/0.5=0.4 [nmol <sub>CYP</sub> ml <sup>-1</sup> ]			→ Eq. 2, 15.7 as [E]	
[329]	Rabbit	7.7	Room T	CYP2B4	Not reported	1.1 [nmol <sub>CYP2B4</sub> ml <sup>-1</sup> ]	k <sub>cat</sub>	nmol min <sup>-1</sup> nmol <sup>-1</sup> <sub>CYP2B4</sub>	→ Eq. 2, 15.9 as [E] <sup>40</sup>	5
[330]	Rabbit	7.4	24.5	CYP2B4	Not reported	1.0 [nmol <sub>CYP2B4</sub> ml <sup>-1</sup> ]	k <sub>cat</sub>	nmol min <sup>-1</sup> nmol <sup>-1</sup> <sub>CYP2B4</sub>	→ Eq. 2, 15.9 as [E] (see previous note)	4
[331] <sup>41</sup>	Rabbit	7.4	25	CYP2B4	Not reported	0.1 [nmol <sub>CYP2B4</sub> ml <sup>-1</sup> ]	k <sub>cat</sub>	min <sup>-1</sup>	→ Eq. 2, 15.9 as [E] (see previous note)	8

<sup>38</sup> In this paper [59], values of some isoenzymes' abundance are given for rabbit

Isoenz.	Isoenz. original paper	Enzyme abund [E]	Source
CYP2B4	P450 LM2 (PB ind.)	<b>16.4</b> nmol <sub>CYP2B4</sub> mg <sup>-1</sup> <sub>PROT</sub>	Haugen DA and Coon MJ (1976) J Biol Chem 251, 7929-7939
CYP1A2	P450 LM4 (BF ind.)	13.8 nmol <sub>CYP1A2</sub> mg <sup>-1</sup> <sub>PROT</sub>	Haugen DA and Coon MJ (1976) J Biol Chem 251, 7929-7939
CYP2E1	P450 LM3a (E ind.)	16-20 nmol <sub>CYP2E1</sub> mg <sup>-1</sup> <sub>PROT</sub>	Koop DR, Morgan ET, Tarr G and Coon MJ (1982) J Biol Chem 257, 8472-8480

<sup>39</sup> This value (18.8) is from sources [60-61] and it's in the interval 16-20 reported in the paper.

<sup>40</sup> This value (15.9±0.4) is the average value from sources [59-61] (see footnote to ref. [54]).

<sup>41</sup> Division of V<sub>max</sub> by the enzyme concentration yields the turnover number or apparent first-order rate constant k<sub>cat</sub>, expressed in reciprocal minutes.





# **Appendix C**

## **Appendix to Chapter 3**

**Table C1.** List of dithiocarbamate (DTC), carbamate (CM) and organophosphorous (OP) pesticides present in the FMO database. The general structures and the ECOSAR classes are also reported.

Group	General structure	Compound class (ECOSAR)	Compound name(s)
Pesticides (DTCs)	$\begin{array}{c} \text{S} \\ \parallel \\ \text{R}^2 - \text{N} - \text{C} - \text{S} - \text{R}^1 \\   \\ \text{R}^3 \end{array}$	Thiocarbamate, Di(Substit) + Thiocarbamate, Di (Na salt)	metam-sodium sodium diethyl-dithiocarbamate sodium dimethyl-dithiocarbamate
		Aliphatic Amines + Thiocarbamate, Di(Substit)	dazomet
		Vinyl/Allyl Halides + Thiocarbamate, Di(Substit)	CDEC (sulfallate)
Pesticides (CMs)	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}^2 - \text{N} - \text{C} - \text{O} - \text{R}^1 \\   \\ \text{R}^3 \end{array}$	Oxime Carbamate Ester	aldicarb thiofanox
		Carbamate Esters	ethiofencarb methiocarb
		Esters + Esters (phosphate)	demeton-S demeton-S-methyl fosthietan phorate oxon
Pesticides (OP)	$\begin{array}{c} \text{X} \\ \parallel \\ \text{R}^1 - \text{X} - \text{P} - \text{X} - \text{R}^3 \\   \\ \text{X} - \text{R}^2 \end{array}$ <p>(X can be either O or S)</p>	Esters + Esters (phosphate) + Nearest analog analysis: pesticides	S-phenyl diethylphosphinothiolothionate
		Nearest analog analysis: pesticides + Esters, Dithiophosphates	disulfoton fonofos phorate sulprofos terbufos
		Nearest analog analysis: pesticides + Esters, Monothiophosphates	demeton-O fenthion

**Table C2.** Relationships between Log  $K_{ow}$  and Log ( $1/K_m$ ) for ADH, including also their ranges and 95% CI of slope and intercept.

Name	Slope( $\pm$ SE)	95%CI slope	Intercept( $\pm$ SE)	95%CI interc.	n	R <sup>2</sup>	SE	p <sup>a</sup>	p <sub>ancova</sub> <sup>b</sup>	Log $K_{ow}$ range	Log ( $1/K_m$ ) range
Regression made merging all species (mammals) and all isoenzymes											
ADHgen	0.59( $\pm$ 0.09)	0.40; 0.78	-3.36( $\pm$ 0.18)	-3.73; -3.00	34	0.56	0.82	<0.01	/	-2.23; 5.50	-5.75; -1.16
Regressions made for the separate species (mammals) and the separate isoenzymes											
ADH1_hor	0.40( $\pm$ 0.11)	0.18; 0.63	-3.08( $\pm$ 0.24)	-3.58; -2.59	20	0.45	0.72	<0.01	0.96	-0.34; 5.50	-4.15; -1.08
ADH1_hum	0.58( $\pm$ 0.12)	0.32; 0.83	-3.01( $\pm$ 0.23)	-3.48; -2.54	24	0.50	0.85	<0.01	0.13	-1.36; 5.50	-4.66; 0.00
ADH2_hum	0.67( $\pm$ 0.19)	0.26; 1.07	-3.58( $\pm$ 0.41)	-4.45; -2.71	18	0.43	1.43	<0.01	0.70	-2.23; 5.50	-5.75; -0.83
ADH3_hum	0.54( $\pm$ 0.25)	-0.11; 1.19	-4.38( $\pm$ 0.58)	-5.87; -2.89	7	0.48	0.72	<u>0.09</u>	<u>&lt;0.01</u>	0.42; 3.53	-4.47; -1.89
ADH1_rat	0.62( $\pm$ 0.19)	0.21; 1.03	-3.11( $\pm$ 0.32)	-3.82; -2.40	13	0.50	0.84	0.01	0.28	-0.77; 3.53	-5.58; -1.11
ADH3_rat	1.18( $\pm$ 0.32)	0.29; 2.07	-6.57( $\pm$ 0.75)	-8.66; -4.48	6	0.77	0.82	0.02	<u>&lt;0.01</u>	0.63; 3.53	-6.28; -2.00

<sup>a</sup> The underlined values indicate non significant regression ( $p>0.05$ ); <sup>b</sup> the underlined values indicate regression significantly different from ADHgen ( $p_{ancova}<0.05$ ).

**Table C3.** Relationships between Log  $K_{ow}$  and Log ( $1/K_m$ ) for ALDH, including also their ranges and 95% CI of slope and intercept, together with 3 additional general regressions leaving out the possibly influential data: I) rat data; II) substituted benzaldehydes; III) rat data as well as substituted benzaldehydes.

Name	Slope(±SE)	95%CI slope	Intercept(±SE)	95%CI interc.	n	R <sup>2</sup>	SE	p <sup>a</sup>	p <sub>ancova</sub> <sup>b</sup>	Log $K_{ow}$ range	Log ( $1/K_m$ ) range
Regression made merging all species (mammals) and all isoenzymes											
ALDHgen	0.69(±0.11)	0.46; 0.92	-1.18(±0.22)	-1.63; -0.73	77	0.33	1.33	<0.01	/	-3.18; 4.19	-2.51; 3.15
Regressions made for the separate species (mammals) and the separate isoenzymes											
ALDH1_hor	0.99(±0.30)	0.29; 1.69	-1.31(±0.38)	-2.18; -0.44	10	0.57	1.00	0.01	0.84	-1.63; 1.90	-2.97; 1.00
ALDH2_hor	0.73(±0.35)	-0.09; 1.56	-0.43(±0.43)	-1.46; 0.59	9	0.39	1.13	0.07	0.10	-1.63; 1.90	-2.43; 1.00
ALDH1_hum	0.82(±0.08)	0.65; 0.98	-0.99(±0.17)	-1.34; -0.64	28	0.80	0.73	<0.01	0.19	-3.18; 3.78	-2.52; 2.60
ALDH2_hum	0.86(±0.13)	0.59; 1.13	-0.73(±0.27)	-1.26; -0.19	57	0.42	1.17	<0.01	<0.01	-1.63; 4.19	-2.71; 3.40
ALDH3_hum	0.54(±0.17)	0.17; 0.92	-1.18(±0.21)	-1.66; -0.71	12	0.51	0.74	0.01	0.95	-3.18; 1.78	-2.94; 0.22
ALDH1_rat	0.18(±0.10)	-0.02; 0.38	-1.33(±0.17)	-1.67; -0.98	32	0.10	0.73	0.08	<0.01	-1.66; 3.76	-1.91; 1.00
ALDH2_rat	0.10(±0.17)	-0.25; 0.45	-2.34(±0.26)	-2.88; -1.80	22	0.02	1.00	0.55	<0.01	-1.66; 2.60	-3.81; -0.60
ALDH3_rat	0.56(±0.33)	-0.25; 1.38	-3.80(±0.74)	-5.62; -1.98	8	0.32	0.45	0.14	<0.01	1.48; 2.88	-3.06; -1.80
I. Regression made merging all species (mammals) and all isoenzymes, excluding 22 substituted benzaldehydes											
	0.81(±0.09)	0.64; 0.99	-1.15(±0.17)	-1.49; -0.81	55	0.63	0.96	<0.01	/	-3.18; 4.19	-2.46; 3.15
II. Regression made merging horse and human data and all isoenzymes											
	0.83(±0.10)	0.63; 1.03	-0.84(±0.20)	-1.24; -0.44	63	0.53	1.05	<0.01	/	-3.18; 4.19	-2.61; 3.15
III. Regression made merging horse and human data and all isoenzymes, excluding substituted benzaldehydes											
	0.83(±0.09)	0.64; 1.01	-0.92(±0.19)	-1.30; -0.54	50	0.63	0.96	<0.01	/	-3.18; 4.19	-2.61; 3.15

<sup>a</sup> The underlined values indicate non significant regression ( $p>0.05$ ); <sup>b</sup> the underlined values indicate regression significantly different from ALDHgen ( $p_{ancova}<0.05$ ).

**Table C4.** Relationships between Log  $K_{ow}$  and Log (1/ $K_m$ ) for FMO, including also their ranges and 95% CI of slope and intercept, together with an additional regression developed including OP pesticides only.

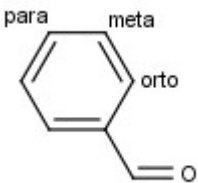
Name	Slope(±SE)	95%CI slope	Intercept(±SE)	95%CI interc.	n	R <sup>2</sup>	SE	p	P <sub>ancova</sub>	Log $K_{ow}$ range	Log (1/ $K_m$ ) range
Regression made merging all species (mammals) and all isoenzymes											
FMOgen	0.22(±0.04)	0.15; 0.29	-2.52(±0.11)	-2.74; -2.29	149	0.20	0.88	<0.01	/	-2.62; 7.49	-4.60; -0.21
Regressions made for the separate species (mammals)											
FMO_mou	0.21(±0.06)	0.09; 0.33	-2.24(±0.16)	-2.56; -1.92	45	0.23	0.80	<0.01	0.08	-2.62; 5.90	-3.46; -0.29
FMO_pig	0.21(±0.04)	0.14; 0.29	-2.48(±0.12)	-2.71; -2.24	144	0.18	0.90	<0.01	0.80	-2.62; 7.49	-4.60; -0.04
Regression made for OP pesticides, merging all species (mammals) and all isoenzymes											
	0.32(±0.09)	0.11; 0.52	-2.34(±0.33)	-3.07; -1.62	12	0.54	0.45	0.01	/	0.68; 5.48	-2.51; -0.21

**Table C5.** Relationships between Log  $K_{ow}$  and Log (1/ $K_m$ ) for CYP, including also their ranges and 95% CI of slope and intercept, together with 5 additional general regressions for separate ECOSAR classes: I) Anilines (Aromatic Amines); II) Benzyl Alcohols; III) Esters; IV) Amides/Imides; V) 'remaining chemicals'.

Name	Slope(±SE)	95%CI slope	Intercept(±SE)	95%CI interc.	n	R <sup>2</sup>	SE	p <sup>a</sup>	P <sub>anova</sub>	Log $K_{ow}$ range	Log (1/ $K_m$ ) range
Regression made merging all species (mammals) and all isoenzymes											
CYPgen	0.34(±0.08)	0.18; 0.50	-3.38(±0.17)	-3.72; -3.05	121	0.13	0.82	<0.01	/	-0.77; 3.71	-4.71; -0.37
b) Regressions made for the separate species (mammals) and the separate isoenzymes											
CYP1A1_rat	0.52(±0.17)	0.16; 0.87	-3.63(±0.32)	-4.29; -2.97	23	0.30	0.54	0.01	0.75	0.46; 3.33	-4.23; -1.85
CYP2B1_rat	0.08(±0.21)	-0.34; 0.50	-2.55(±0.48)	-3.51; -1.59	39	0.00	1.02	<u>0.70</u>	0.09	-0.07; 3.63	-4.71; -0.37
CYP2B4_rab	0.24(±0.12)	0.00; 0.48	-3.39(±0.27)	-3.93; -2.85	47	0.08	0.76	<u>0.05</u>	0.12	0.23; 3.71	-4.57; -1.49
CYP2E1_rab	0.78(±0.10)	0.58; 0.98	-4.00(±0.16)	-4.32; -3.68	36	0.65	0.51	<0.01	0.94	-0.77; 3.32	-4.70; -1.60
I. Regression made for Anilines (Aromatic Amines) merging all the species (mammals) and all the isoenzymes											
	0.77(±0.26)	0.21; 1.33	-4.19(±0.46)	-5.16; -3.21	17	0.37	0.51	0.01	/	0.90; 2.86	-4.23; -2.13
II. Regression made for Benzyl Alcohols merging all the species (mammals) and all the isoenzymes											
	0.84(±0.20)	0.41; 1.27	-4.03(±0.32)	-4.71; -3.35	17	0.54	0.37	<0.01	/	0.62; 2.38	-3.83; -1.78
III. Regression made for Esters merging all the species (mammals) and all the isoenzymes											
	0.84(±0.14)	0.54; 1.14	-4.48(±0.26)	-5.03; -3.94	17	0.70	0.54	<0.01	/	0.03; 3.32	-4.70; -1.58
IV. Regression made for Amides/imides merging all the species (mammals) and all the isoenzymes											
	0.48(±0.13)	0.20; 0.76	-3.03(±0.23)	-3.53; -2.52	14	0.54	0.43	<0.01	/	-0.07; 3.33	-3.23; -1.28
V. Regression made for the remaining chemicals merging all the species (mammals) and all the isoenzymes											
	0.16(±0.13)	-0.10; 0.43	-3.02(±0.33)	-3.68; -2.36	56	0.03	0.99	<u>0.22</u>	/	-0.77; 3.71	-4.71; -0.37

<sup>a</sup> The underlined values indicate non significant regression (p>0.05)

**Table C6.** List of the 22 substituted benzaldehydes present in ALDH database; their general structure is also reported with the positions of the substituents.

General structure of substituted benzaldehydes	Compound class (ECOSAR)	Compound name <sup>a</sup>
	<i>Aldehydes (Mono) + Dinitrobenz.</i>	<u>2,4-dinitrobenzaldehyde</u>
	<i>Aldehydes (Mono)</i>	3,4-dimethoxybenzaldehyde
	<i>Aldehydes (Mono) + Phenols</i>	m-hydroxybenzaldehyde
	<i>Aldehydes (Mono)</i>	m-methoxybenzaldehyde
	<i>Aldehydes (Mono)</i>	m-methylbenzaldehyde
	<i>Aldehydes (Mono)</i>	<u>o-bromobenzaldehyde</u>
	<i>Aldehydes (Mono)</i>	o-chlorobenzaldehyde
	<i>Aldehydes (Mono)</i>	<u>o-fluorobenzaldehyde</u>
	<i>Aldehydes (Mono) + Phenols</i>	<u>o-hydroxybenzaldehyde</u>
	<i>Aldehydes (Mono)</i>	o-methoxybenzaldehyde
	<i>Aldehydes (Mono)</i>	o-methylbenzaldehyde
	<i>Aldehydes (Mono)</i>	<u>o-nitrobenzaldehyde</u>
	<i>Aldehydes (Mono)</i>	p-(dimethylamino)-benzaldehyde
	<i>Aldehydes (Mono)</i>	<u>p-bromobenzaldehyde</u>
	<i>Aldehydes (Mono)</i>	p-carboxybenzaldehyde
	<i>Aldehydes (Mono)</i>	<u>p-chlorobenzaldehyde</u>
	<i>Aldehydes (Mono)</i>	p-cyanobenzaldehyde
	<i>Aldehydes (Mono)</i>	<u>p-fluorobenzaldehyde</u>
	<i>Aldehydes (Mono)</i>	<u>p-iodobenzaldehyde</u>
	<i>Aldehydes (Mono)</i>	p-methoxybenzaldehyde
	<i>Aldehydes (Mono)</i>	p-methylbenzaldehyde
	<i>Aldehydes (Mono)</i>	p-nitrobenzaldehyde

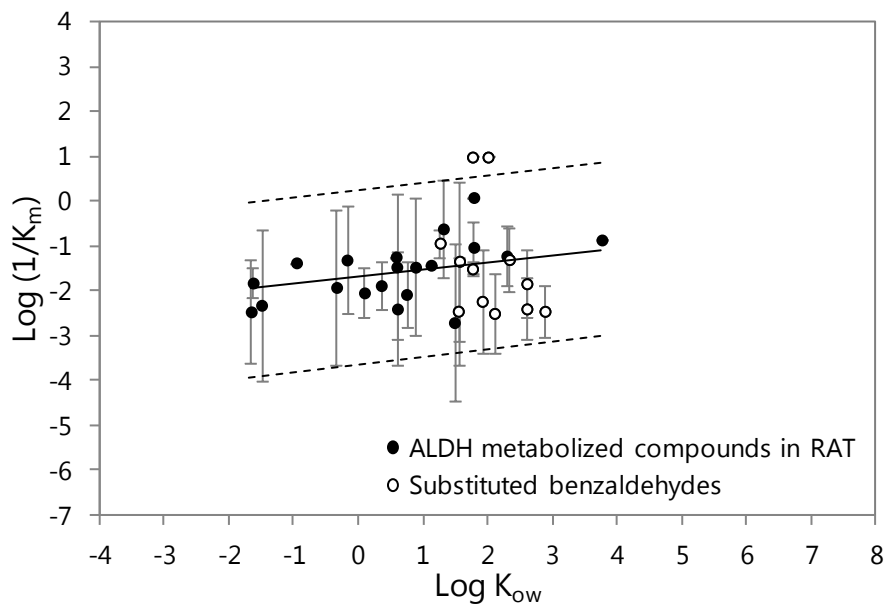
<sup>a</sup> The underlined compounds are outliers of ALDHgen regression.

**Table C7.** Relationships between Log  $K_{ow}$  and Log ( $1/K_m$ ) for ALDH in rat (with and without substituted benzaldehydes), including also their ranges and 95% CI of slope and intercept. The  $K_m$  values were expressed as  $\mu\text{M}$ .

Slope ( $\pm$ SE)	95%CI slope	Intercept ( $\pm$ SE)	95%CI interc.	n	$r^2$	SE	$p^a$	Log $K_{ow}$ range	Log( $1/K_m$ ) range
Regression including all compounds									
0.16 ( $\pm$ 0.12)	-0.09; 0.41	-1.69 ( $\pm$ 0.21)	-2.12; -1.26	32	0.06	0.91	<u>0.19</u>	-1.66; 3.76	-2.70; 0.10
Regression excluding substituted benzaldehydes									
0.26 ( $\pm$ 0.10)	0.05; 0.47	-1.71 ( $\pm$ 0.15)	-2.02; -1.41	20	0.28	0.60	0.02	-1.66; 3.76	-2.70; 1.00

<sup>a</sup> The underlined values indicate non significant regression ( $p>0.05$ )

**Figure C1.** Relationship between Log  $K_{ow}$  and Log ( $1/K_m$ ) in rat for compounds metabolised by ALDH. Regressions (solid lines) and 95% confidence intervals (dashed lines). Laboratory measurements (dots): Log transformed geometrical mean of  $1/K_m$  [ $\mu\text{M}^{-1}$ ] for each compound, with the geometric standard deviation (vertical bar). White dots correspond to substituted benzaldehydes.





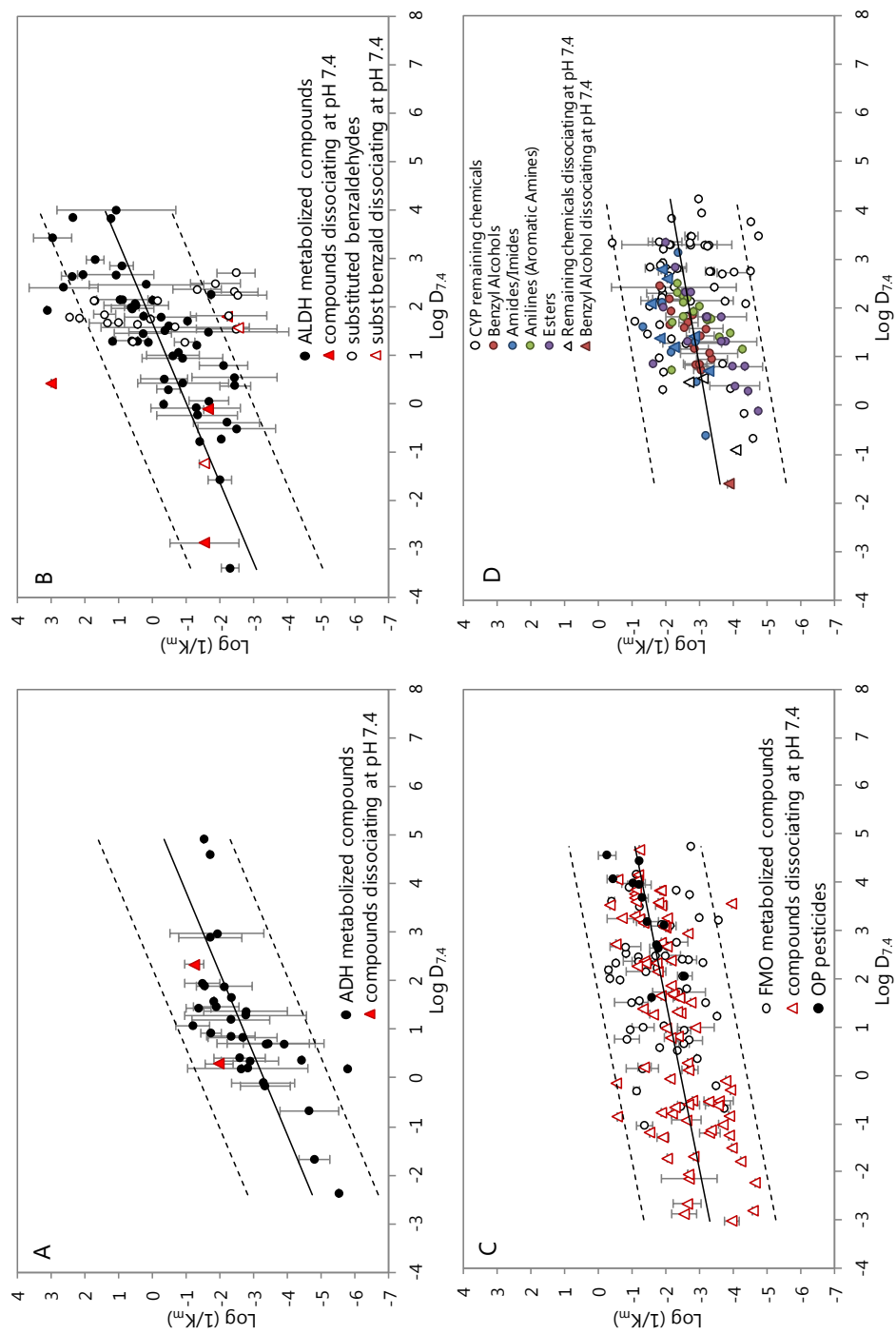
Regressions considering ionisation (ionis), obtained merging all species and isoenzymes for the four enzymes families.

**Table C8.** Relationships between  $\text{Log } D_{7.4}$  and  $\text{Log } (1/K_m)$ , including also their ranges and 95% CI of slope and intercept, and the percentage of compounds with a dissociated fraction larger than 0.05 at pH 7.4. As a comparison, the corresponding relationships obtained with  $\text{Log } K_{ow}$  are reported. The  $K_m$  values were expressed as  $\mu\text{M}$ .

Name	Slope( $\pm\text{SE}$ )	95%CI slope	Intercept( $\pm\text{SE}$ )	95%CI interc.	n	r <sup>2</sup>	SE	p	Log D range <sup>a</sup>	Log(1/ $K_m$ ) range	% ionised compounds
ADHgen ionis	0.60( $\pm 0.10$ )	0.39; 0.81	-3.30( $\pm 0.18$ )	-3.67; -2.93	34	0.52	0.85	<0.01	-2.39; 4.90	-5.75; -1.16	5.9%
ADHgen	0.59( $\pm 0.09$ )	0.40; 0.78	-3.36( $\pm 0.18$ )	-3.73; -3.00	34	0.56	0.82	<0.01	-2.23; 5.50	-5.75; -1.16	
ALDHgen ionis	0.61( $\pm 0.12$ )	0.37; 0.84	-1.00( $\pm 0.23$ )	-1.45; -0.55	77	0.26	1.40	<0.01	-3.43; 3.97	-2.51; 3.15	7.8%
ALDHgen	0.69( $\pm 0.11$ )	0.46; 0.92	-1.18( $\pm 0.22$ )	-1.63; -0.73	77	0.33	1.33	<0.01	-3.18; 4.19	-2.51; 3.15	
FMOgen ionis	0.29( $\pm 0.04$ )	0.22; 0.36	-2.43( $\pm 0.09$ )	-2.60; -2.26	148 <sup>b</sup>	0.31	0.82	<0.01	-3.03; 4.73	-4.60; -0.21	54.1%
FMOgen	0.22( $\pm 0.04$ )	0.15; 0.29	-2.52( $\pm 0.11$ )	-2.74; -2.29	149	0.20	0.88	<0.01	-2.62; 7.49	-4.60; -0.21	
CYPgen ionis	0.25( $\pm 0.07$ )	0.11; 0.39	-3.20( $\pm 0.15$ )	-3.50; -2.90	121	0.10	0.83	<0.01	-1.62; 4.22	-4.71; -0.37	9.1%
CYPgen	0.34( $\pm 0.08$ )	0.18; 0.50	-3.38( $\pm 0.17$ )	-3.72; -3.05	121	0.13	0.82	<0.01	-0.77; 3.71	-4.71; -0.37	

<sup>a</sup> D= $D_{7.4}$  in the regressions considering ionisation, otherwise D= $K_{ow}$ ; <sup>b</sup> the  $\text{Log } D_{7.4}$  value of one compound (2-aminoazulene) was not available.

**Figure C2** (next page). Relationships between  $\text{Log } D_{7.4}$  and  $\text{Log } (1/K_m)$  in mammals for compounds metabolised by: A) ADH; B) ALDH; C) FMO; D) CYP. Regressions (solid lines) and 95% confidence intervals (dashed lines). Laboratory measurements (dots):  $\text{Log transformed geometrical mean of } 1/K_m [\mu\text{M}^{-1}]$  for each compound, with the geometric standard deviation (vertical bar). The compounds with a dissociated fraction larger than 0.05 at pH 7.4 are represented with triangles.



**Table C9.** Regressions developed with the sets of data used in the reviews by Lewis et al. 2004 [30] and Hansch et al. 2004 [26], recalculated in this study using only Log K<sub>ow</sub> as descriptor (experimental value, when available) and K<sub>m</sub> expressed in μM.

# <sup>a</sup>	Isoenz.	Species	Slope	95%CI slope	Intercept	95%CI interc	n	r <sup>2</sup>	SE	p <sup>b</sup>	LogK <sub>ow</sub> range	Log (1/K <sub>m</sub> ) range	Ref
<u>S8.1</u>	CYP2B1	Rat	0.90(±0.15)	0.56; 1.23	-3.72(±0.31)	-4.40; -3.04	14	0.74	0.54	<0.01	-0.07; 3.35	-4.03; -0.37	[325]
<u>S8.2</u>	CYP2E1	Rabbit	0.77(±0.13)	0.49; 1.06	-4.30(±0.23)	-4.80; -3.79	15	0.72	0.49	<0.01	0.03; 3.32	-4.70; -1.60	[327]
<u>S8.3</u>	CYP2B4	Rabbit	0.87(±0.14)	0.51; 1.23	-5.11(±0.29)	-5.86; -4.36	7	0.88	0.39	<0.01	0.23; 3.32	-4.57; -2.30	[327]
<u>S8.4</u>	CYP1A1	Rat	1.09(±0.14)	0.79; 1.39	-4.89(±0.25)	-5.43; -4.35	15	0.83	0.25	<0.01	0.90; 2.86	-4.23; -2.30	[31]
<u>S8.5</u>	CYP2B4	Rabbit	0.73(±0.10)	0.51; 0.94	-5.25(±0.28)	-5.88; -4.62	13	0.83	0.17	<0.01	2.09; 3.42	-3.72; -2.58	[322]
<u>S8.6</u>	CYP2E1	Rabbit	0.66(±0.23)	0.17; 1.15	-3.53(±0.36)	-4.31; -2.76	17	0.36	0.42	0.01	0.62; 2.38	-3.88; -1.70	[328]
<u>S8.7</u>	CYP2B4	Rabbit	1.01(±0.22)	0.54; 1.49	-4.52(±0.35)	-5.28; -3.77	17	0.58	0.41	<0.01	0.62; 2.38	-3.86; -1.78	[328]
S8.8	CYP2B1	Rat	0.34(±0.32)	-0.41; 1.09	-2.86(±0.80)	-4.75; -0.97	9	0.14	0.36	<u>0.32</u>	1.81; 3.06	-2.57; -1.48	[324]
S8.9	CYP2E1	Rabbit	0.56(±0.05)	0.33; 0.79	-4.09(±0.03)	-4.23; -3.95	4	0.98	0.07	0.01	-0.77; 0.88	-4.54; -3.64	[326]
S8.10	CYP1A1	Rat	0.25(±0.13)	-0.07; 0.57	-2.68(±0.26)	-3.30; -2.05	8	0.38	0.34	<u>0.11</u>	0.46; 3.33	-2.86; -1.85	[321]
S8.11	CYP2B4	Rabbit	-0.95(±0.68)	-2.60; 0.70	-0.10(±2.01)	-5.01; 4.80	8	0.25	1.01	<u>0.21</u>	2.09; 3.71	-4.34; -0.90	[331]
S8.12	CYP2B4	Rabbit	-0.41(±0.84)	-4.04; 3.22	-1.70(±0.69)	-4.66; 1.26	4	0.11	0.42	<u>0.67</u>	0.51; 1.07	-2.56; -1.76	[330]
S8.13	CYP2B4	Rabbit	-0.56(±0.26)	-1.39; 0.26	-0.27(±0.74)	-2.63; 2.09	5	0.61	0.18	<u>0.12</u>	2.41; 3.24	-2.13; -1.49	[329]
S8.14	CYP2B1	Rat	-0.61(±0.33)	-2.73; 0.72	-1.00(±0.83)	-1.29; 0.06	25	0.13	1.00	<u>0.07</u>	1.29; 3.63	-4.71; -1.04	[323]

<sup>a</sup> The underlined regressions are the ones considered acceptable (n>6, p<0.05); <sup>b</sup> the underlined values indicate non significant regression (p>0.05)



# **Appendix D**

## **Appendix to Chapter 4**

**Table D1.** Log (1/*K<sub>m</sub>*): Variables selected and their non-standardised regression coefficients (± error). The *K<sub>m</sub>* values were expressed as μM.

Enzyme	logP	A	a/d <sup>2</sup>	1/w	apK <sub>a</sub> 1	bpK <sub>a</sub> 1	HBD	HBA	v	E <sub>HOMO</sub>	E <sub>LUMO</sub>	ΔE <sub>L-H</sub>	H <sub>i</sub>	Interc.	ALDH	FMO	CYP
ADH	0.66 (±0.11)		-6.4E-10 <sup>a</sup> (±3.5E-10)		-0.08 (±0.03)		-0.35 <sup>a</sup> (±0.20)				-0.19 (±0.07)			-2.57 (±0.28)	/	/	/
ALDH	0.50 (±0.15)	1.3E-2 (±3.0E-3)					-0.52 (±0.25)	0.57 (±0.21)				0.38 (±0.16)	8.0E-3 (±3.0E-3)	-7.78 (±2.01)	/	/	/
FMO		1.9E-3 (±5.8E-4)						-0.45 (±0.07)				-0.19 (±0.05)		-0.46 <sup>a</sup> (±0.50)	/	/	/
CYP		5.2E-3 (±1.2E-3)		0.65 (±0.22)	-0.06 (±0.02)				-0.06 <sup>a</sup> (±0.04)			-0.23 (±0.05)		-2.43 (±0.64)	/	/	/
ALL	0.29 (±0.04)		-5.9E-10 <sup>a</sup> (±3.2E-10)	0.67 (±0.18)						-0.11 <sup>a</sup> (±0.06)		-0.11 (±0.05)		-2.31 (±0.62)	1.83 (±0.22)	-0.28 <sup>a</sup> (±0.21)	-0.84 (±0.20)
Additional regressions																	
ALDH <sub>1</sub>	0.45 (±0.10)	8.8E-3 (±2.0E-3)		0.85 <sup>a</sup> (±0.44)	-0.14 (±0.04)	0.10 (±0.05)	-0.44 (±0.18)							-3.44 (±0.64)	/	/	/
ALDH <sub>2</sub>	-1.42 <sup>a</sup> (±0.84)	4.3E-2 (±1.3E-2)										1.48 <sup>a</sup> (±0.88)		-19.36 (±8.18)	/	/	/
CYP <sub>1</sub>	0.44 (±0.09)	6.1E-3 (±1.1E-3)			-0.03 <sup>a</sup> (±0.02)		-0.15 <sup>a</sup> (±0.10)				-0.33 (±0.09)		2.2E-3 (±1.1E-3)	-4.42 (±0.21)	/	/	/
CYP <sub>2</sub>				2.04 (±0.39)	-0.04 <sup>a</sup> (±0.03)							-0.21 (±0.07)		-3.80 (±0.99)	/	/	/

<sup>a</sup> The probability (p) value of the coefficient is greater than 0.05.

**Table D2.** Log  $V_{\max}$ : Variables selected and their non-standardised regression coefficients ( $\pm$  error). The  $V_{\max}$  values were expressed as  $\mu\text{mol}\cdot\text{min}^{-1}\cdot\text{mg}_{\text{PROT}}^{-1}$ .

Enzyme	logP	A	a/d <sup>2</sup>	I/w	apK <sub>a</sub> 1	bpK <sub>a</sub> 1	HBD	HBA	v	E <sub>HOMO</sub>	E <sub>LUMO</sub>	$\Delta E_{\text{L-H}}$	H <sub>f</sub>	Interc.
ADH								-0.30 ( $\pm 0.10$ )	0.26 ( $\pm 0.07$ )	-0.23 ( $\pm 0.04$ )				0.68 ( $\pm 0.17$ )
ALDH	-0.28 ( $\pm 0.07$ )	5.9E-3 ( $\pm 1.7\text{E-}3$ )			0.04 <sup>a</sup> ( $\pm 0.03$ )				-0.05 <sup>a</sup> ( $\pm 0.03$ )	0.24 ( $\pm 0.09$ )			3.3E-3 <sup>a</sup> ( $\pm 1.7\text{E-}3$ )	-0.93 ( $\pm 0.25$ )
FMO				-0.28 ( $\pm 0.11$ )	-0.04 ( $\pm 0.01$ )	0.02 <sup>a</sup> ( $\pm 0.01$ )		0.13 ( $\pm 0.03$ )	0.02 <sup>a</sup> ( $\pm 0.01$ )	0.13 ( $\pm 0.04$ )				1.14 ( $\pm 0.40$ )
CYP			1.4E-3 ( $\pm 5.5\text{E-}4$ )		0.06 ( $\pm 0.02$ )	-0.14 <sup>a</sup> ( $\pm 0.10$ )			-0.10 ( $\pm 0.03$ )	0.21 ( $\pm 0.06$ )			3.7E-3 ( $\pm 1.1\text{E-}3$ )	-1.22 ( $\pm 0.11$ )
<b>Additional regressions</b>														
ALDH <sub>1</sub>	-0.19 ( $\pm 0.06$ )	3.5E-3 ( $\pm 1.4\text{E-}3$ )		0.54 <sup>a</sup> ( $\pm 0.28$ )		-0.26 ( $\pm 0.11$ )			-0.07 ( $\pm 0.02$ )					-1.23 ( $\pm 0.43$ )
ALDH <sub>2</sub>		1.6E-2 ( $\pm 6.6\text{E-}3$ )				1.17 ( $\pm 0.53$ )								-3.95 ( $\pm 1.31$ )
CYP <sub>1</sub>	-0.26 ( $\pm 0.10$ )		2.6E-3 ( $\pm 6.6\text{E-}4$ )		0.06 ( $\pm 0.03$ )			0.34 ( $\pm 0.15$ )	-0.24 ( $\pm 0.06$ )	0.32 ( $\pm 0.14$ )				-1.54 ( $\pm 0.30$ )
CYP <sub>2</sub>		4.5E-3 ( $\pm 1.6\text{E-}3$ )		0.71 ( $\pm 0.24$ )	-0.04 <sup>a</sup> ( $\pm 0.02$ )	0.28 <sup>a</sup> ( $\pm 0.15$ )		-0.34 ( $\pm 0.09$ )				0.13 ( $\pm 0.06$ )		-4.11 ( $\pm 0.93$ )

<sup>a</sup> The probability (p) value of the coefficient is greater than 0.05.

**Table D3.** Applicability domains for Log (1/K<sub>m</sub>) QSARs.

Enzyme	logP	A	a/d <sup>2</sup>	I/w	apK <sub>a</sub> 1	bpK <sub>a</sub> 1	HBD	HBA	v	E <sub>HOMO</sub>	E <sub>LUMO</sub>	ΔE <sub>L-H</sub>	H <sub>f</sub>
ADH	-2.22; 4.82		1.4E+0; 2.1E+9		0; 23		0; 3				-1.40; 6.49		
ALDH	-4.65; 3.66	53.8; 424.0					0; 3	1; 5				5.00; 11.58	-162.38; 130.63
FMO		80.8; 652.0						0; 4				4.87; 12.81	
CYP		71.2; 429.0		1.00; 2.42	0; 19			0.02; 13.14				7.72; 14.91	
ALL	-4.65; 6.66		1.4E+0; 2.1E+9	1.00; 2.45						-4.27; 6.49		4.87; 14.91	
<b>Additional regressions</b>													
ALDH <sub>1</sub>	-4.65; 3.66	53.8; 424.0		1.02; 2.45	0; 10	0; 15	0; 3						
ALDH <sub>2</sub>	1.34; 2.61	168.0; 251.0										7.70; 9.29	
CYP <sub>1</sub>	-0.55; 3.71	101.0; 401.0			0; 12		0; 2			-1.20; 3.63		-154.58; 36.96	
CYP <sub>2</sub>				1.00; 2.11	0; 19							7.72; 14.91	



**Table D4.** Applicability domains for Log V<sub>max</sub> QSARs.

Enzyme	logP	A	a/d <sup>2</sup>	I/w	apK <sub>a</sub> 1	bpK <sub>a</sub> 1	HBD	HBA	v	E <sub>HOMO</sub>	E <sub>LUMO</sub>	ΔE <sub>L-H</sub>	H <sub>f</sub>
ADH								1; 4	0.61; 7.61		-1.40; 6.49		
ALDH	-4.65; 3.66	53.8; 424.0			0; 11				0.01; 28.58		-4.27; 2.65		-162.38; 130.63
FMO				1.00; 2.32	0; 18	0; 10		0; 4	0.50; 22.80	-12.23; - 4.62			
CYP			1.37; 468.77		0; 19		0; 2		0.02; 13.14	-1.20; 3.78			-154.58; 61.07
<b>Additional regressions</b>													
ALDH <sub>1</sub>	-4.65; 3.66	53.8; 424.0		1.02; 2.45			0; 3		0.01; 28.58				
ALDH <sub>2</sub>		168.0; 251.0					0; 1						
CYP <sub>1</sub>	-0.55; 3.71		1.57; 468.77			0; 12		1; 3	0.72; 13.14	-1.20; 3.63			
CYP <sub>2</sub>		71.2; 429.0		1.00; 2.11		0; 9	0; 2	0; 3				7.72; 14.91	

**Table D5.** Relationships between  $\text{Log } K_{ow}$  and  $\text{Log } (1/K_m)$  for the four enzyme groups (from [79]), merging all species (mammals) and all isoenzymes. The  $K_m$  values were expressed as  $\mu\text{M}$ .

Name	Slope( $\pm\text{SE}$ )	95%CI slope	Intercept( $\pm\text{SE}$ )	95%CI interc.	N	$r^2$	RMSE	p	$Q^2_{\text{Loo}}$	RMSE <sub>Loo</sub>	$\text{Log } K_{ow}$ range	$\text{Log } (1/K_m)$ range
ADHgen	0.59( $\pm 0.09$ )	0.40; 0.78	-3.36( $\pm 0.18$ )	-3.73; -3.00	34	0.56	0.79	<0.01	0.50	0.85	-2.23; 5.50	-5.75; -1.16
ALDHgen	0.69( $\pm 0.11$ )	0.46; 0.92	-1.18( $\pm 0.22$ )	-1.63; -0.73	77	0.33	1.31	<0.01	0.30	1.34	-3.18; 4.19	-2.51; 3.15
FMOgen	0.22( $\pm 0.04$ )	0.15; 0.29	-2.52( $\pm 0.11$ )	-2.74; -2.29	149	0.20	0.72	<0.01	0.18	0.89	-2.62; 7.49	-4.60; -0.21
CYPgen	0.34( $\pm 0.08$ )	0.18; 0.50	-3.38( $\pm 0.17$ )	-3.72; -3.05	121	0.13	0.81	<0.01	0.10	0.82	-0.77; 3.71	-4.71; -0.37

**Table D6.** Relationships between  $\text{Log } K_{ow}$  and  $\text{Log } (1/K_m)$  for specific groups of chemicals for ALDH and CYP. The  $K_m$  values were expressed as  $\mu\text{M}$ .

Name	Slope( $\pm\text{SE}$ )	95%CI slope	Intercept( $\pm\text{SE}$ )	95%CI interc.	n	$r^2$	RMSE	$p^a$	$Q^2_{\text{Loo}}$	RMSE <sub>Loo</sub>	$\text{Log } K_{ow}$ range	$\text{Log } (1/K_m)$ range
<b>ALDH</b>												
1. Regression made merging all species (mammals) and all isoenzymes, excluding the 22 substituted benzaldehydes												
	0.81( $\pm 0.09$ )	0.64; 0.99	-1.15( $\pm 0.17$ )	-1.49; -0.81	55	0.63	0.95	<0.01	0.60	0.98	-3.18; 4.19	-2.46; 3.15
2. Regression made for the 22 substituted benzaldehydes merging all species (mammals) and all isoenzymes												
	-1.35( $\pm 0.79$ )	-2.99; 0.29	2.22( $\pm 1.54$ )	-0.98; 5.43	22	0.13	1.59	<u>0.10</u>	0.02	1.72	1.22; 2.88	-2.51; 2.49
<b>CYP</b>												
1. Regression made merging all species (mammals) and all isoenzymes, excluding the 'remaining chemicals'												
	-3.91( $\pm 0.17$ )	-4.25; -3.56	0.70( $\pm 0.10$ )	0.50; 0.90	65	0.45	0.56	<0.01	0.41	0.58	-0.07; 3.33	-4.70; -1.28
2. Regression made for the 'remaining chemicals' merging all species (mammals) and all isoenzymes												
	0.16( $\pm 0.13$ )	-0.10; 0.43	-3.02( $\pm 0.33$ )	-3.68; -2.36	56	0.03	0.97	<u>0.22</u>	0.01	1.01	-0.77; 3.71	-4.71; -0.37

<sup>a</sup> The underlined values indicate non significant regression ( $p > 0.05$ ).

# Appendix **E**

## Appendix to Chapter 5

**Table E1.** Applicability domains (AD) for Log (1/K<sub>m</sub>) QSARs, defined by the range (min and max) of the values of the descriptors for the data in the training sets. For each enzyme, also the range (min and max) of Log (1/K<sub>m</sub>) values are reported.

Enzyme	Name	Group	AD training set	AD test set
ADH	nOHs	Dragon6	(0; 2)	(0; 1)
	SIC4	Dragon6	(0; 0.97)	(0; 0.84)
	Mor23u	Dragon6	(-2.59; 0.04)	(-1.79; 0.14)
	Log (1/K <sub>m</sub> )		(-5.75; -1.16)	(-4.76; -1.44)
ALDH	3DACorr_PiChg_2	Adriana	(-0.07; 0.16)	(-0.04; 0)
	MAT55v	Dragon6	(-0.30; 0.76)	(-0.25; 0.74)
	Mor01e	Dragon6	(17.7; 751)	(6.6; 440)
	XLogP	CDK	(-0.53; 6.83)	(-0.70; 4.57)
	InertiaY	Adriana	(9.03; 1680)	(1.76; 1760)
	Log (1/K <sub>m</sub> )		(-2.51; 3.15)	(-2.49; 3)
FMO	RHSA	CDK	(0.53; 1)	(0.58; 1)
	Se1N1N2ss	E-state	(0; 6.31)	(0; 6.13)
	N-067	Dragon6	(0; 2)	(0; 2)
	Hy	Dragon6	(-0.93; 3.20)	(-0.94; 5)
	2DACorr_LpEN_1	Adriana	(2.38; 285)	(2.57; 146)
	R4e+	Dragon6	(0; 0.18)	(0; 0.17)
	Log (1/K <sub>m</sub> )		(-4.60; -0.21)	(-4.54; -0.26)
CYP	AROM	Dragon6	(0; 1)	(0; 1)
	ATS7v	Dragon6	(0; 3.15)	(0; 3.19)
	PDI	Dragon6	(0.62; 0.997)	(0.72; 0.95)
	RTu	Dragon6	(7.1; 30.3)	(8.8; 24.6)
	JGI5	Dragon6	(0; 0.08)	(0; 0.08)
	C2SP3	CDK	(0; 9)	(0; 8)
	Log (1/K <sub>m</sub> )		(-4.71; -0.37)	(-4.70; -1.28)

**Table E2.** Applicability domains (AD) for Log  $V_{\max}$  QSARs, defined by the range (min and max) of the values of the descriptors for the data in the training sets. For each enzyme, also the range (min and max) of Log  $V_{\max}$  values are reported.

Enzyme	Name	Group	AD training set	AD test set
ADH	nHDon	Dragon6	(0; 3)	(0; 1)
	tautomercount	Chemaxon	(1; 2)	(1; 2)
	Mor15s	Dragon6	(-2.62; 5.73)	(-0.87; 0.80)
	ASP	Dragon6	(0.16; 0.97)	(0.27; 0.90)
	Log $V_{\max}$		(-0.32; 1.93)	(-0.70; 1.45)
ALDH	nArX	Dragon6	(0; 1)	(0; 1)
	R6m+	Dragon6	(0; 0.32)	(0; 0.50)
	Mor26e	Dragon6	(-0.26; 0.42)	(-0.23; 0.45)
	WNSA-1	CDK	(37.8; 231)	(26.4; 163)
	Log $V_{\max}$		(-2.50; 0.84)	(-1.64; 0.63)
FMO	Se1C3N3as	EState	(0; 5.84)	(0; 6.3)
	Se2C3O1s	E-state	(0; 11.9)	(0; 9.78)
	Log $V_{\max}$		(-1.31; 0.40)	(-1.10; 0.36)
CYP	Se1C1C3sd	E-state	(0; 2.01)	(0; 2.07)
	Mor24s	Dragon6	(-1.56; 3.40)	(-1.17; 1.12)
	Mor10s	Dragon6	(-4.64; 6.18)	(-3.37; 2.42)
	formalcharge_pH_7.4	Chemaxon	(-1; 1)	(-1; 1)
	Log $V_{\max}$		(-3.18; -0.24)	(-3.11; -0.42)



# Appendix **F**

## Appendix to Chapter 6

**Table F1.** Clearance  $CL_{INT}$  ( $\mu\text{L}/\text{min}/10^6$  cells) measured *in vitro* in human hepatocytes for pharmaceuticals and environmental pollutants, together with their name, CAS number, SMILES, type (PPP = plant protection product) and dataset (i.e., training or test set). Information on how clearance was measured (SD = substrate depletion, PF = product formation) is also reported.

Name	CAS n	Type	$CL_{INT}$ (HEP) ( $\mu\text{L}/\text{min}/10^6$ cells)	Source	Comment	Dataset
abamectin	65195-55-3	PPP (insecticide)	13.10	[332]	SD	Not incl.
acetochlor	34256-82-1	PPP (herbicide)	66.00	[333]	SD	Training
aflatoxin	1162-65-8	Natural product (mycotoxin)	3.49	[332]	SD	Training
alachlor	15972-60-8	PPP (herbicide)	50.90	[333]	SD	Test
aminogluthetimide	125-84-8	Pharmaceutical	0.72	[332]	SD	Training
amitriptyline	50-48-6	Pharmaceutical	1.17	[332]; [146]	SD, Average of 2 studies (st dev $\pm 0.30$ )	Training
amphetamine	300-62-9	Pharmaceutical	0.51	[332]	SD	Training
atenolol	29122-68-7	Pharmaceutical	1.90	[146]	SD	Training
atrazine	1912-24-9	PPP (herbicide)	3.34	[333]	SD	Training
atropine	51-55-8	Natural product (alkaloid)	0.50	[332]	SD	Test
bensulide	741-58-2	PPP (herbicide)	169.00	[333]	SD	Training
bisphenol a	80-05-7	Industrial Chemical	22.20	[333]	SD	Training
bufuralol	54340-62-4	Pharmaceutical	5.10	[146]	SD	Training
buprofezin	69327-76-0	PPP (insecticide)	15.00	[333]	SD	Test
busulfan	55-98-1	Pharmaceutical	1.39	[332]	SD	Training
caffeine	58-08-2	Natural product (alkaloid)	0.11	[332]	SD	Training
carbamazepine	298-46-4	Pharmaceutical	0.26	[332]	SD	Test
carbaryl	63-25-2	PPP (insecticide)	7.37	[332]	SD	Test
carvedilol	72956-09-3	Pharmaceutical	29.00	[146]	SD	Test



cerivastatin	145599-86-6	Pharmaceutical	6.10	[146]	SD	Training
chloramphenicol	56-75-7	Pharmaceutical	0.72	[332]	SD	Training
chlorpheniramine	132-22-9	Pharmaceutical	1.00	[146]	SD	Test
chlorpromazine	50-53-3	Pharmaceutical	17.00	[146]	SD	Test
chlorpyrifos	2921-88-2	PPP (insecticide)	2.60	[332]	SD	Training
cimetidine	51481-61-9	Pharmaceutical	13.00	[146]	SD	Training
clothianidin	210880-92-5	PPP (insecticide)	10.50	[333]	SD	Training
clozapine	5786-21-0	Pharmaceutical	7.60	[146]	SD	Training
colchicine	64-86-8	Natural product	0.70	[332]	SD	Test
cycloheximide	66-81-9	Natural product	1.74	[332]	SD	Training
cyprodinil	121552-61-2	PPP (fungicide)	19.30	[333]	SD	Test
DDT	50-29-3	PPP (insecticide)	1.43	[334]	SD	Training
desipramine	50-47-5	Pharmaceutical	3.30	[146]	SD	Training
dexamethasone	50-02-2	Pharmaceutical	1.60	[146]	SD	Test
dextropropoxyphene	469-62-5	Pharmaceutical	11.00	[332]	SD	Test
diazepam	439-14-5	Pharmaceutical	0.43	[332]; [146]	SD, Average of 2 studies (st dev $\pm 0.77$ )	Training
diazoxon	962-58-3	PPP (insecticide)	57.30	[333]	SD	Training
dibutylphthalate	84-74-2	Industrial Chemical	42.50	[332]	SD	Training
diclofenac	15307-86-5	Pharmaceutical	19.00	[146]	SD	Training
dicrotophos	141-66-2	PPP (insecticide)	0.94	[333]	SD	Training
diethylphthalate	84-66-2	Industrial Chemical	42.50	[332]	SD	Training
diltiazem	42399-41-7	Pharmaceutical	6.80	[146]	SD	Training
diphenhydramine	58-73-1	Pharmaceutical	0.89	[332]; [146]	SD, Average of 2 studies (st dev $\pm 0.49$ )	Training
diphenylhydantoin	57-41-0	Pharmaceutical	0.72	[332]	SD	Test
disopyramide	3737-09-5	Pharmaceutical	0.46	[332]	SD	Training

diuron	330-54-1	PPP (herbicide)	6.07	[332], [333]*	SD	Test
dofetilide	115256-11-6	Pharmaceutical	1.90	[146]	SD	Test
etoxazole	153233-91-1	PPP (pesticide)	17.00	[333]	SD	Training
fenamiphos	22224-92-6	PPP (insecticide)	46.20	[333]	SD	Training
fenoxycarb	72490-01-8	PPP (insecticide)	13.80	[333]	SD	Training
fenpropathrin	39515-41-8	PPP (pesticide)	3.17	[332]	SD	Test
fenvaletrate	51630-58-1	PPP (insecticide)	0.77	[332]	SD	Test
fipronil	120068-37-3	PPP (insecticide)	0.16	[332]	SD	Training
forchlorfenuron	68157-60-8	PPP (pesticide)	13.40	[333]	SD	Test
gemfibrozil	25812-30-0	Pharmaceutical	16.00	[146]	SD	Training
glipizide	29094-61-9	Pharmaceutical	1.00	[146]	SD	Training
haloperidol	52-86-8	Pharmaceutical	0.99	[332]	SD	Training
hydrocortisone	50-23-7	Pharmaceutical	8.10	[146]	SD	Training
ibuprofen	15687-27-1	Pharmaceutical	4.71	[332]; [146]	SD, Averag of 2 studies (st dev ±0.22)	Test
imipramine	50-49-7	Pharmaceutical	6.60	[146]	SD	Training
isoniazide	54-85-3	Pharmaceutical	1.11	[332]	SD	Test
isoxaben	82558-50-7	PPP (herbicide)	2.06	[333]	SD	Training
isoxaflutole	141112-29-0	PPP (pesticide)	23.60	[333]	SD	Test
ketoprofen	22071-15-4	Pharmaceutical	3.70	[146]	SD	Training
lidocaine	137-58-6	Pharmaceutical	6.20	[146]	SD	Training
lorcainide	59729-31-6	Pharmaceutical	5.90	[146]	SD	Training
malathion	121-75-5	PPP (insecticide)	42.50	[332]	SD	Test
maprotiline	10262-69-8	Pharmaceutical	0.72	[332]	SD	Training
metalaxyl	57837-19-1	PPP (fungicide)	4.32	[332]	SD	Training
methadone	76-99-3	Pharmaceutical	1.34	[332]	SD	Training

methylparathion	298-00-0	PPP (insecticide)	10.50	[333]	SD	Training
methylphenidate	113-45-1	Pharmaceutical	2.51	[332]	SD	Test
methylprednisolone	83-43-2	Pharmaceutical	2.20	[146]	SD	Test
metoprolol	37350-58-6	Pharmaceutical	1.40	[146]	SD	Test
metribuzin	21087-64-9	PPP (herbicide)	0.54	[333]	SD	Training
MGK	113-48-4	PPP (pesticide)	76.60	[333]	SD	Test
midazolam	59467-70-8	Pharmaceutical	14.00	[146]	SD	Training
nadolol	42200-33-9	Pharmaceutical	3.00	[146]	SD	Training
nalidixic acid	389-08-2	Pharmaceutical	1.10	[332]	SD	Training
naloxone	465-65-6	Pharmaceutical	25.00	[146]	SD	Training
naproxen	22204-53-1	Pharmaceutical	2.00	[146]	SD	Test
nicotine	54-11-5	Natural product (alkaloid)	2.34	[332]	SD	Training
nifedipine	21829-25-4	Pharmaceutical	13.00	[146]	SD	Training
omeprazole	73590-58-6	Pharmaceutical	3.00	[146]	SD	Training
ondansetron	99614-02-5	Pharmaceutical	1.00	[146]	SD	Test
orphenadrine	83-98-7	Pharmaceutical	0.72	[332]	SD	Training
paracetamol (acetaminophen)	103-90-2	Pharmaceutical	0.85	[332]; [146]	SD, Average of 2 studies (st dev $\pm 0.10$ )	Training
parathion	56-38-2	PPP (insecticide)	3.32	[332], [333]*	SD	Test
PCB136 (2,2',3,3',6,6'- Hexachlorobiphenyl)	38411-22-2	Industrial Chemical	2.7E-04	[335]	SD	Test
PCB153 (2,2',4,4',5,5'- Hexachlorobiphenyl)	35065-27-1	Industrial Chemical	2.7E-06	[335]	SD	Training
PCB155 (2,2',4,4',6,6'- Hexachlorobiphenyl)	33979-03-2	Industrial Chemical	9.2E-06	[335]	SD	Test
PCB77 (3,3',4,4'- Tetrachlorobiphenyl)	32598-13-3	Industrial Chemical	6.8E-05	[335]	SD	Training

PCB80 (3,3',5,5'- Tetrachlorobiphenyl)	33284-52-5	Industrial Chemical	3.3E-05	[335]	SD	Training
pentobarbital	76-74-4	Pharmaceutical	0.72	[332]	SD	Test
phenacetin	62-44-2	Pharmaceutical	6.30	[146]	SD	Test
physostigmine	57-47-6	Natural product (alkaloid)	1.64	[332]	SD	Training
prazosin	19216-56-9	Pharmaceutical	4.30	[146]	SD	Test
prednisolone	50-24-8	Pharmaceutical	4.90	[146]	SD	Training
procainamide	51-06-9	Pharmaceutical	0.72	[332]	SD	Training
propafenone	54063-53-5	Pharmaceutical	25.00	[146]	SD	Training
propanolol (propranolol)	525-66-6	Pharmaceutical	2.88	[332]; [146]	SD, Average of 2 studies (st dev $\pm 0.61$ )	Test
propetamphos	31218-83-4	PPP (insecticide)	9.77	[333]	SD	Training
propylparaben	94-13-3	Industrial Chemical (preservative)	42.50	[332]	SD	Training
pyraclostrobin	175013-18-0	PPP (fungicide)	35.00	[333]	SD	Training
pyrithiobac	123343-16-8	PPP (herbicide)	1.99	[333]	SD	Training
quinidine	56-54-2	Pharmaceutical	2.23	[332]; [146]	SD, Average of 2 studies (st dev $\pm 0.43$ )	Training
risperidone	106266-06-2	Pharmaceutical	9.70	[146]	SD	Test
rotenone	83-79-4	PPP (insecticide)	22.70	[332], [333]*	SD	Training
sildenafil	139755-83-2	Pharmaceutical	8.10	[146]	SD	Test
strychnine	57-24-9	PPP (insecticide)	2.00	[332]	SD	Training
tenoxicam	59804-37-4	Pharmaceutical	1.00	[146]	SD	Training
thiazopyr	117718-60-2	PPP (pesticide)	41.20	[333]	SD	Test
thioridazine	50-52-2	Pharmaceutical	0.92	[332]	SD	Test
tolbutamide	64-77-7	Pharmaceutical	1.10	[146]	SD	Training
triadimefon	43121-43-3	PPP (fungicide)	16.30	[333]	SD	Training

triclosan	3380-34-5	Industrial Chemical (antibacterial)	83.80	[333]	SD	Training
verapamil	52-53-9	Pharmaceutical	8.89	[332]; [146]	SD, Average of 2 studies (st dev $\pm 0.23$ )	Training
warfarin	81-81-2	Pharmaceutical	0.72	[332]	SD	Training
zolpidem	82626-48-0	Pharmaceutical	2.80	[146]	SD	Training
zoxamide	156052-68-5	PPP (fungicide)	28.80	[333]	SD	Training

\* The value measured by Rotroff et al. 2010 [333] was used, as in Tonnelier et al. 2012 [145].

**Table F2.** Clearance  $CL_{INT}$  (L/min/mg<sub>MICR</sub>) measured *in vitro* in human microsomes for pharmaceuticals and environmental pollutants, together with their name, CAS number, SMILES, type (PPP = Plant protection product) and dataset (i.e., training or test set). Information on how clearance was measured is also reported (SD = substrate depletion, PF = product formation).

Name	CAS n	Type	$CL_{INT}$ (MICR) ( $\mu$ L/min/ mg <sub>MICR</sub> )	Source	Comment	Dataset
1,2,3,3,3-pentafluoropropene	5528-43-8	Industrial Chemical	560.78	[336]	PF, Sum of 3 metabolites	Training
1,2,3,4-diepoxybutane	30419-67-1	Industrial Chemical	32.46	[337]	PF, Hydroxylated metabolites	Test
1,2-dibromoethane	106-93-4	Industrial Chemical	4.62	[338]	PF, One metabolite	Training
1,3-butadiene	106-99-0	Industrial Chemical	113.66	[339]; [340]	PF, SD, Average of 2 studies (st dev $\pm 0.43$ ), same metabolite	Training
2,2',3,3',6,6'-hexachlorobiphenyl (236HCB)	38411-22-2	Industrial Chemical	0.58	[341]	PF, Hydroxylated metabolites	Training
2,2',4,4',5-pentabromodiphenyl ether (BDE-99)	60348-60-9	Industrial Chemical	32.85	[342]	PF, Sum of 6 metabolites	Training

2,2',4,4'- tetrabromodiphenyl ether (BDE-47)	5436-43-1	Industrial Chemical	7.56	[343]; [344]	PF, SD, Average of 2 studies (st dev $\pm 0.51$ ), sum of 4 and 3 metabolites each	Training
2-ethylhexyl tetrabromobenzoate	183658-27-7	Industrial Chemical	58.02	[345]	PF, One metabolite	Test
4,4'-dichlorobiphenyl	2050-68-2	Industrial Chemical	2.79	[346]	PF, Hydroxylated metabolites	Training
4-chlorophenyl methyl sulphide	123-09-1	Industrial Chemical	8.29	[307]	PF, One metabolite	Test
8-2 fluorotelomer alcohol	678-39-7	Industrial Chemical	15.59	[347]	SD	Test
8-hydroxy-2,3,7-trichloro- dibenzo-p-dioxin	NA	Industrial Chemical	32.26	[348]	PF, One metabolite	Training
alprazolam	28981-97-7	Pharmaceutical	1.78	[349]	SD, oxidised metabolites	Training
amitriptyline	50-48-6	Pharmaceutical	16.26	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.03$ )	Training
amobarbital	57-43-2	Pharmaceutical	1.04	[349]	SD, oxidised metabolites	Test
atenolol	29122-68-7	Pharmaceutical	14.00	[146]	SD, oxidised metabolites	Training
benfuracarb	82560-54-1	PPP (insecticide)	204.25	[350]; [351]	PF, SD, Average of 2 studies (st dev $\pm 0.25$ ), sum of same metabolites	Test
benzene	71-43-2	Industrial Chemical	27.97	[352]	PF, Oxidised metabolites	Test
benzo[a]pyrene	50-32-8	Industrial Chemical	34.48	[353]	SD	Training
beta-chloroprene	126-99-8	Industrial Chemical	1666.67	[354]	PF, Oxidised metabolites	Training
betaxolol	63659-18-7	Pharmaceutical	2.60	[146]	SD, oxidised metabolites	Test
bisphenol A	80-05-7	Industrial Chemical	703.89	[355]; [356]	PF, SD, Average of 2 studies (st dev $\pm 0.03$ ), same metabolite	Training
bisphenol F (4,4'- dihydroxydiphenyl- methane)	620-92-8	Industrial Chemical	1.71	[357]	PF, One metabolite	Training
bosentan	147536-97-8	Pharmaceutical	14.00	[146]	SD, oxidised metabolites	Test
bromobenzene	108-86-1	Industrial Chemical	14.14	[358]	PF, Sum of 2 metabolites	Training

bufuralol	54340-62-4	Pharmaceutical	18.00	[146]	SD, oxidised metabolites	Training
butadiene monoxide	930-22-3	Industrial Chemical	28.18	[359]; [360]	PF, Sum of different metabolites (oxidised and hydroxylated)	Training
carbaryl	63-25-2	PPP (insecticide)	9.64	[361]	PF, Sum of 3 metabolites	Training
carbofuran	1563-66-2	PPP (insecticide)	1.67	[362]	PF, One metabolite	Test
carbosulfan	55285-14-8	PPP (insecticide)	534.41	[363]	PF, Sum of 3 metabolites	Test
carvedilol	72956-09-3	Pharmaceutical	167.00	[146]	SD, oxidised metabolites	Test
cerivastatin	145599-86-6	Pharmaceutical	22.00	[146]	SD, oxidised metabolites	Training
chlorobenzene	108-90-7	Industrial Chemical	13.04	[358]	PF, Sum of 2 metabolites	Training
chlorpromazine	50-53-3	Pharmaceutical	40.14	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.23$ )	Training
chlorpyrifos	2921-88-2	PPP (insecticide)	54.35	[364]; [365]; [366]; [367]	PF, SD, Average of 4 studies (st dev $\pm 0.46$ ), sum of same 2 metabolites	Test
cimetidine	51481-61-9	Pharmaceutical	10.00	[146]	SD, oxidised metabolites	Training
clozapine	5786-21-0	Pharmaceutical	9.04	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.35$ )	Training
deltamethrin	52918-63-5	PPP (insecticide)	120.14	[368]	SD	Test
desipramine	50-47-5	Pharmaceutical	20.84	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.06$ )	Training
dexamethasone	50-02-2	Pharmaceutical	5.77	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.34$ )	Training
diazepam	439-14-5	Pharmaceutical	4.07	[161]; [349]; [146]	SD, Average of 3 studies (st dev $\pm 0.41$ )	Training
diazinon	333-41-5	PPP (pesticide)	60.58	[369]; [365]	PF, SD, Average of 2 studies (st dev $\pm 0.02$ ), sum of same 2 metabolites	Training
dibenzo[def,p]chrysene	191-30-0	Industrial Chemical	9.84	[353]	SD	Training
diclofenac	15307-86-5	Pharmaceutical	208.49	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.004$ )	Training
diltiazem	42399-41-7	Pharmaceutical	36.51	[161]; [349]; [146]	SD, Average of 3 studies (st dev $\pm 0.38$ )	Test
dimethylformamide	68-12-2	Industrial Chemical	4.75	[370]	PF, One metabolite	Test

diphenhydramine	58-73-1	Pharmaceutical	4.73	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.43$ )	Training
diphenyl sulphide	139-66-2	Industrial Chemical	10.80	[307]	PF, One metabolite	Test
disulfoton	298-04-4	PPP (pesticide)	133.33	[371]	PF, One metabolite	Training
diuron	330-54-1	PPP (herbicide)	174.24	[372]	PF, One metabolite	Training
endosulfan- $\alpha$	115-29-7	PPP (pesticide)	18.21	[373]	PF, One metabolite	Test
esfenvalerate	66230-04-4	PPP (insecticide)	27.64	[368]	SD	Training
ethyl methyl sulphide	624-89-5	Industrial Chemical	0.73	[307]	PF, One metabolite	Training
fenthion	55-38-9	PPP (pesticide)	51.10	[242]	PF, Sum of 2 metabolites	Training
fipronil	120068-37-3	PPP (insecticide)	3.95	[374]	PF, One metabolite	Training
FK1052	129300-27-2	Pharmaceutical	46.96	[161]	SD, oxidised metabolites	Test
FK480	150408-73-4	Pharmaceutical	59.17	[161]	SD, oxidised metabolites	Training
furosemide	54-31-9	Pharmaceutical	30.00	[146]	SD, oxidised metabolites	Test
gemfibrozil	25812-30-0	Pharmaceutical	45.00	[146]	SD, oxidised metabolites	Training
hexachloro-1,3-butadiene	87-68-3	Industrial Chemical	1.56	[375]	PF, One metabolite	Training
hexobarbital	56-29-1	Pharmaceutical	2.56	[349]	SD, oxidised metabolites	Training
hydrocortisone	50-23-7	Pharmaceutical	49.00	[146]	SD, oxidised metabolites	Training
ibuprofen	15687-27-1	Pharmaceutical	18.50	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.39$ )	Training
imipramine	50-49-7	Pharmaceutical	25.99	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.13$ )	Training
ketamine	6740-88-1	Pharmaceutical	30.00	[349]	SD, oxidised metabolites	Training
ketoprofen	22071-15-4	Pharmaceutical	8.10	[146]	SD, oxidised metabolites	Test
lorazepam	846-49-1	Pharmaceutical	23.00	[146]	SD, oxidised metabolites	Test
lorcainide	59729-31-6	Pharmaceutical	103.28	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.38$ )	Training
methiocarb	2032-65-7	PPP (pesticide)	59.42	[371]	PF, One metabolite	Training
methohexital	151-83-7	Pharmaceutical	54.44	[349]	SD, oxidised metabolites	Training
methoxsalen	298-81-7	Pharmaceutical	44.44	[349]	SD, oxidised metabolites	Training



methyl parathion	298-00-0	PPP (pesticide)	138.66	[369]	PF, Sum of 2 metabolites	Training
methyl tertiary-butyl ether	1634-04-4	Industrial Chemical	8.53	[376]	PF, One metabolite	Training
methylene chloride	75-09-2	Industrial Chemical	3.15	[377]	PF, One metabolite	Training
methylethyl ketoxime	96-29-7	Industrial Chemical	0.51	[378]	PF, One metabolite	Test
methylprednisolone	83-43-2	Pharmaceutical	28.00	[146]	SD, oxidised metabolites	Training
metoprolol	37350-58-6	Pharmaceutical	4.30	[146]	SD, oxidised metabolites	Test
midazolam	59467-70-8	Pharmaceutical	250.51	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.21$ )	Training
molinate	2212-67-1	PPP (herbicide)	3.56	[379]	PF, Sum of 2 metabolites	Test
N-(hydroxymethyl)-N-methylformamide	NA	Industrial Chemical	0.03	[370]	PF, One metabolite	Training
nadolol	42200-33-9	Pharmaceutical	20.00	[146]	SD, oxidised metabolites	Training
naloxone	465-65-6	Pharmaceutical	14.00	[146]	SD, oxidised metabolites	Training
naphthalene	91-20-3	Industrial Chemical	131.76	[380]	PF, Sum of 3 metabolites	Training
naproxen	22204-53-1	Pharmaceutical	21.00	[146]	SD, oxidised metabolites	Training
nicardipine	55985-32-5	Pharmaceutical	1384.46	[161]	SD, oxidised metabolites	Training
nifedipine	21829-25-4	Pharmaceutical	96.00	[146]	SD, oxidised metabolites	Test
nilvadipine	75530-68-6	Pharmaceutical	1365.76	[161]	SD, oxidised metabolites	Test
N-methylformamide (NMF)	123-39-7	Industrial Chemical	0.06	[370]	PF, One metabolite	Training
nonane	111-84-2	Industrial Chemical	6888.87	[274]	PF, Sum of 2 metabolites	Test
omeprazole	73590-58-6	Pharmaceutical	37.52	[161]; [146]	SD, Average of 2 studies (st dev $\pm 0.45$ )	Training
paracetamol (acetaminophen)	103-90-2	Pharmaceutical	8.10	[146]	SD, oxidised metabolites	Training
parathion	56-38-2	PPP (pesticide)	68.06	[364]	PF, Sum of 2 metabolites	Training
perfluorooctanesulfonamide (FOSA)	754-91-6	Industrial Chemical	2.75	[381]	PF, One metabolite	Training

phenacetin	62-44-2	Pharmaceutical	27.00	[146]	SD, oxidised metabolites	Test
phorate	298-02-2	PPP (pesticide)	135.78	[371]	PF, One metabolite	Test
prazosin	19216-56-9	Pharmaceutical	8.10	[146]	SD, oxidised metabolites	Training
prednisone	53-03-2	Pharmaceutical	3.00	[349]	SD, oxidised metabolites	Test
profenofos	41198-08-7	PPP (pesticide)	9286.07	[382]; [383]	PF, SD, Average of 2 studies (st dev $\pm 0.68$ ), sum of 2 and 1 metabolites each	Training
propafenone	54063-53-5	Pharmaceutical	189.16	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.02$ )	Training
propranolol (propranolol)	525-66-6	Pharmaceutical	22.00	[146]	SD, oxidised metabolites	Training
propylene oxide	75-56-9	Industrial Chemical	27.39	[384]	PF, Hydroxylated metabolites	Training
quinidine	56-54-2	Pharmaceutical	9.72	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.58$ )	Training
risperidone	106266-06-2	Pharmaceutical	43.00	[146]	SD, oxidised metabolites	Test
sildenafil	139755-83-2	Pharmaceutical	60.00	[146]	SD, oxidised metabolites	Test
styrene	100-42-5	Industrial Chemical	23.33	[385]	PF, One metabolite	Training
sulprofos	35400-43-2	PPP (pesticide)	20.39	[371]	PF, One metabolite	Test
tenidap	120210-48-2	Pharmaceutical	9.22	[349]	SD, oxidised metabolites	Test
tenoxicam	59804-37-4	Pharmaceutical	1.89	[349]	SD, oxidised metabolites	Test
theophylline	58-55-9	Pharmaceutical	3.10	[146]	SD, oxidised metabolites	Training
tolbutamide	64-77-7	Pharmaceutical	1.45	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.23$ )	Training
triazolam	28911-01-5	Pharmaceutical	21.11	[349]	SD, oxidised metabolites	Test
trichloroethylene	79-01-6	Industrial Chemical	56.15	[386]	PF, One metabolite	Training
verapamil	52-53-9	Pharmaceutical	138.74	[349]; [146]	SD, Average of 2 studies (st dev $\pm 0.01$ )	Training
warfarin	81-81-2	Pharmaceutical	2.20	[146]	SD, oxidised metabolites	Training
zolpidem	82626-48-0	Pharmaceutical	9.78	[161]; [349]; [146]	SD, Average of 3 studies (st dev $\pm 0.45$ )	Test

**Table F3.** Applicability domains (AD) for Log CL<sub>INT</sub> QSARs for human hepatocytes and microsomes, defined by the range (min and max) of the values of the descriptors for the data in the training sets. Also the range (min and max) of Log CL<sub>INT</sub> values are reported.

Assay	Name	Group	AD training set	AD test set
Hepatocytes	2DACorr_SigChg_5	Adriana	(-0.42; 0.71)	(-0.64; 0.43)
	R8u+	Dragon6	(0.00; 0.05)	(0.00; 0.05)
	R5e+	Dragon6	(0.02; 0.13)	(0.02; 0.10)
	2DACorr_SigChg_2	Adriana	(-0.96; 0.03)	(-0.75; -0.02)
	HATS0m	Dragon6	(0.06; 3.35)	(0.06; 3.44)
	Log CL <sub>INT</sub>		(-5.57; 2.23)	(-5.03; 1.88)
Microsomes	smalleststringsize	Chemaxon	(0.00; 6.00)	(0.00; 7.00)
	GATS4v	Dragon6	(0.00; 1.39)	(0.72; 1.39)
	Se2O1P4s	E-States	(0.00; 8.87)	(0.00; 0.00)
	2DACorr_SigChg_9	Adriana	(-0.14; 0.41)	(-0.24; 0.37)
	HATS5e	Dragon6	(0.00; 1.21)	(0.17; 1.19)
	Se2C2O1s	E-States	(0.00; 7.87)	(0.00; 0.00)
	Log CL <sub>INT</sub>		(-1.58; 3.97)	(-0.29; 3.84)



# **Appendix G**

## **Appendix to Chapter 7**

### Tentative comparison between hepatocytes and enzymes

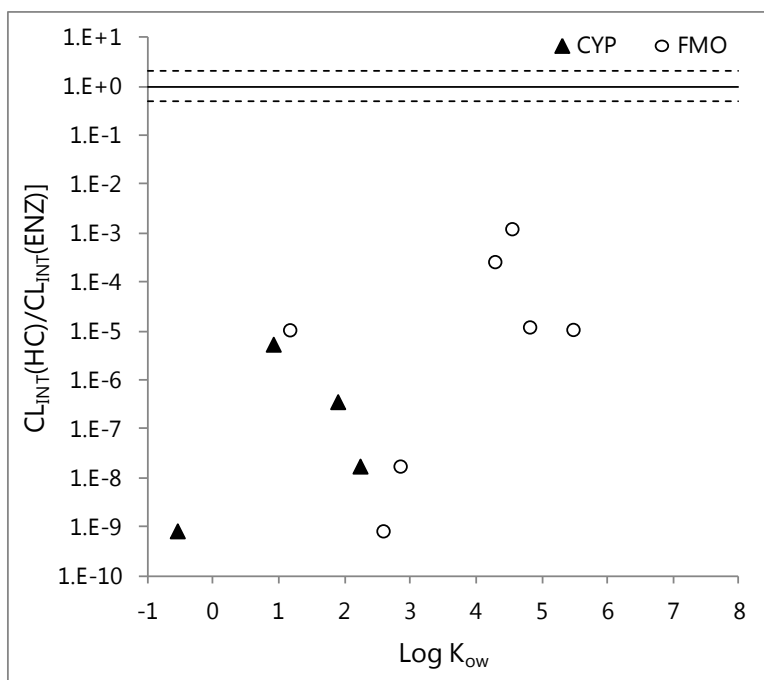
In this synthesis, an empirical equation has been derived to compare the clearance measured in hepatocytes to the clearance measured in enzymes:

$$CL_{\text{hepatocytes}} = f(V_{\text{max,enzyme}}/K_{\text{m,enzyme}}).$$

For this purpose, the datasets were examined to find compounds with data available for hepatocytes and for enzymes. In total, the human hepatocytes dataset had 11 compounds in common with the enzymes datasets (4 chemicals for CYP and 7 for FMO, none of them measured in humans). In order to compare the metabolic constants of hepatocytes to the ones obtained in the enzymatic assays, the clearance values were expressed as intrinsic liver clearances ( $CL_{\text{INT,liver}}$ ,  $\text{L min}^{-1} \text{g}_{\text{LIV}}^{-1}$ ) by multiplying the *in vitro*  $CL_{\text{INT}}$  (i.e. the ratio  $V_{\text{max}}/K_{\text{m}}$ ) by the *in vitro* system scaling factor (SF), as explained in Section 7.3.1. The values of the SFs were taken from Table 7.2 and are  $10^6$  cells  $\text{g}_{\text{LIV}}^{-1}$  for human hepatocytes,  $32 \text{ mg}_{\text{PROT}} \text{g}_{\text{LIV}}^{-1}$  for CYP and  $0.13 \text{ mg}_{\text{PROT}} \text{g}_{\text{LIV}}^{-1}$  for FMO (average of pig and mouse values). As FMO and CYP data are values averaged over different species, the *in vivo* was not performed and the liver  $CL_{\text{INT}}$  values were directly compared for hepatocytes and enzymes data.

For the 11 compounds metabolised by CYP and FMO, the relationship with the hepatocytes data was not statistically relevant, with  $R^2 < 0.1$ . For all compounds, the clearances for hepatocytes were more than 10 fold lower than for enzymes (Figure G1). This can be due to the fact that the enzyme concentrations used for *in vitro* assays are higher than those that are present in the *in vivo* situation, so clearances may be higher than *in vivo*. Given the low statistics ( $r^2 < 0.1$ ), and as hepatocytes were quite able to represent the *in vivo* situation as shown in Section 7.3.2, enzymatic data seem less suitable to perform *in vivo*.

**Figure G1** (next page). Ratio (Log 10 scale) between the *in vivo* liver hepatic clearances  $CL_{\text{INT}}$  ( $\text{L min}^{-1} \text{g}_{\text{LIV}}^{-1}$ ) measured in hepatocytes and in enzymes (CYP: green triangles; FMO: white dots) for 11 compounds for which data were available, in relationship with their  $\text{Log } K_{\text{ow}}$ . The black line represents a ratio of 1 (value in HC = value in MS) and the two dotted lines the 2-fold lower and higher error.



**Table G1.** For ADH, ALDH and FMO enzymes there is no tabulated value for the scaling factors. The content of enzyme in liver was taken from the papers measuring  $K_m$  and  $V_{max}$ , but this value was reported only in few studies. It also varies for the same species and/or isoenzyme. Here are the cases in which the scaling factor was available for ADH, ALDH and FMO:

ADH (out of a total of 13 articles and 199 records)

Ref	# records	Species	Isoenzyme	Protein weight (mg <sub>PROT</sub> g <sub>LIV</sub> <sup>-1</sup> )
[282]	8	Human	ADH1	0.07
[270]	13	Human	ADH2	0.405
[285]	15, 7	Rat	ADH1, ADH3	0.006, 0.017
[286]	12	Human	ADH1	0.322
[287]	13	Horse	ADH1	0.013
[273]	5	Human	ADH3	0.041
Total:	73	All species	Average (±st.d.):	0.12 (±0.17)
		<b>Human</b>	<b>Average (±st.d.):</b>	<b>0.21 (±0.18)</b>

ALDH (out of a total of 17 articles and 244 records)

Ref	# records	Species	Isoenzyme	Protein weight (mg <sub>PROT</sub> g <sub>LIV</sub> <sup>-1</sup> )
[291]	1	Human	ALDH3	0.003
[292]	10, 9	Horse	ALDH1, ALDH2	0.003, 0.002
[294]	2, 2	Human	ALDH1, ALDH2	0.033, 0.083
[297]	6	Human	ALDH3	0.011
[300]	2,2	Human	ALDH1, ALDH2	0.048, 0.170
[303]	15, 14	Rat	ALDH1, ALDH2	0.028, 0.011
Total:	63		Average (±st.d.):	0.04 (±0.05)
		<b>Human</b>	<b>Average (±st.d.):</b>	<b>0.06 (±0.06)</b>

FMO (out of a total of 22 articles and 263 records)

Ref	# records	Species	Isoenzyme	Protein weight (mg <sub>PROT</sub> g <sub>LIV</sub> <sup>-1</sup> )
[313]	15, 17	Mouse, Pig	FMO	0.015, 0.254
Total:	32		Average (±st.d.):	0.13 (±0.17)



**Table G2.** *In vivo* CL<sub>H</sub> (L min<sup>-1</sup> kg<sub>BW</sub><sup>-1</sup>) for human measured from *in vivo* intravenous pharmacokinetics experiments for 22 pharmaceuticals (data from Paixão et al. 2010 [36]), together with the *in vivo* CL<sub>H</sub> (L min<sup>-1</sup> kg<sub>BW</sub><sup>-1</sup>) estimated applying the ivive method described in Section 7.1.1 to *in vitro* CL<sub>INT</sub> data from human hepatocytes (HC) and microsomes (MS). The CL<sub>INT</sub> data used for the ivive are in Table X of Appendix X. The ratio between CL<sub>H</sub> measured *in vivo* and CL<sub>H</sub> extrapolated from HC and MS data are also reported (\*\*if difference greater than 10-fold, \* if difference greater than 2-fold).

Name	Class <sup>a</sup>	logP (ACD)	logD <sub>7.4</sub> (ACD)	CL <sub>H</sub> meas.	CL <sub>H</sub> HC estim.	CL <sub>H</sub> /CL <sub>H</sub> HC	CL <sub>H</sub> MS estim-	CL <sub>H</sub> /CL <sub>H</sub> MS
sildenafil	A	1.65	1.5	0.006	0.010	0.6	0.015	0.4*
diclofenac	A	4.26	1.1	0.007	0.015	0.5	0.019	0.4*
gemfibrozil	A	4.39	1.51	0.001	0.014	0.1**	0.013	0.1**
atenolol	B	0.43	-1.8	0.000	0.004	0.0**	0.007	0.0**
naloxone	B	1.62	0.85	0.018	0.016	1.1	0.007	2.4*
metoprolol	B	1.76	-0.47	0.014	0.003	4.4*	0.003	4.5*
quinidine	B	2.51	0.86	0.004	0.004	1.0	0.006	0.8
propranolol	B	2.58	0.36	0.015	0.005	2.7*	0.010	1.5
risperidone	B	2.63	1.25	0.008	0.011	0.7	0.013	0.6
diltiazem	B	2.73	1.89	0.011	0.009	1.2	0.012	0.9
carvedilol	B	3.42	2.07	0.012	0.016	0.7	0.018	0.7
clozapine	B	3.52	1.1	0.005	0.010	0.5	0.005	1.0
diphenhydramine	B	3.65	2.17	0.006	0.002	2.7*	0.003	1.7
amitriptyline	B	4.81	2.48	0.010	0.003	4.0*	0.008	1.3
cimetidine	N	-0.11	-0.22	0.003	0.013	0.3*	0.006	0.5
paracetamol	N	0.91	0.9	0.002	0.002	1.2	0.005	0.5*
methylprednisolone	N	1.56	1.56	0.004	0.004	0.8	0.011	0.3*
prazosin	N	1.65	1.43	0.004	0.007	0.6	0.005	0.8
nifedipine	N	1.82	1.81	0.004	0.013	0.3*	0.017	0.3*
zolpidem	N	3.02	3.01	0.006	0.005	1.1	0.006	1.0
diazepam	N	3.08	3.08	0.001	0.001	0.6	0.003	0.2*
midazolam	N	3.33	3.28	0.008	0.013	0.6	0.019	0.4*

<sup>a</sup> Compounds were classified as acid (A) if pK<sub>a</sub>1 ≤ 7.4, base (B) if pK<sub>b</sub>1 ≥ 7.4, otherwise neutral (N) (pK calculated with ACD)



## References

### Literature cited

1. Sijm D.T.H.M., et al., 2007. *Transport, accumulation and transformation processes*, in *Risk Assessment of Chemicals*, van Leeuwen C. and Vermeire T. (Eds.). Springer: Dordrecht, The Netherlands. p. 73-158.
2. European Union, 2006. *Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 Concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)*.
3. Sobanska M.A., et al., 2014. *Analysis of the ecotoxicity data submitted within the framework of the REACH Regulation. Part 1. General overview and data availability for the first registration deadline*. *Science of The Total Environment*, 470–471: p. 1225-1232.
4. European Chemicals Agency (ECHA), 2012. *Guidance on registration, available at <http://guidance.echa.europa.eu/>. in Guidance for the implementation of REACH*. Helsinki, Finland.
5. Hendriks A.J., 2013. *How to deal with 100,000+ substances, sites, and species: Overarching principles in environmental risk assessment*. *Environmental Science & Technology*, 47(8): p. 3546-3547.
6. Lilienblum W., et al., 2008. *Alternative methods to safety studies in experimental animals: Role in the risk assessment of chemicals under the new European Chemicals Legislation (REACH)*. *Archives of Toxicology*, 82(4): p. 211-236.
7. Russom C.L., et al., 2003. *An overview of the use of quantitative structure-activity relationships for ranking and prioritizing large chemical inventories for environmental risk assessments*. *Environmental Toxicology and Chemistry*, 22(8): p. 1810-1821.
8. McKinney J.D., et al., 2000. *The practice of Structure Activity Relationships (SAR) in toxicology*. *Toxicological Sciences*, 56(1): p. 8-17.
9. Cronin M.T.D., 2010. *Quantitative Structure-Activity Relationships (QSARs) - Applications and methodology*, in *Recent Advances in QSAR Studies*, Puzyn T., Leszczynski J., and Cronin M.T.D. (Eds.). Springer: Dordrecht, The Netherlands. p. 3-11.
10. Perkins R., et al., 2003. *Quantitative structure-activity relationship methods: Perspectives on drug discovery and toxicology*. *Environmental Toxicology and Chemistry*, 22(8): p. 1666-1679.
11. Hermens J.L.M., et al., 2013. *The octanol–water partition coefficient: Strengths and limitations*. *Environmental Toxicology and Chemistry*, 32(4): p. 732-733.
12. Gramatica P., 2010. *Chemometric methods and theoretical molecular descriptors in predictive QSAR modeling of the environmental behavior of*

*organic pollutants*, in *Recent Advances in QSAR Studies*, Puzyn T., Leszczynski J., and Cronin M.T. (Eds.). Springer: Dordrecht, The Netherlands. p. 327-366.

13. Klaassen C.D., 2008. *Casarett and Doull's Toxicology: The Basic Science of Poisons*. 7th ed. McGraw-Hill: New York (USA).

14. Mackay D. and Fraser A., 2000. *Bioaccumulation of persistent organic chemicals: mechanisms and models*. *Environmental Pollution*, 110(3): p. 375-391.

15. Hendriks A.J., et al., 2001. *The power of size. 1. Rate constants and equilibrium ratios for accumulation of organic substances related to octanol-water partition ratio and species weight*. *Environmental Toxicology and Chemistry*, 20(7): p. 1399-1420.

16. Veltman K., et al., 2009. *Bioaccumulation potential of air contaminants: Combining biological allometry, chemical equilibrium and mass-balances to predict accumulation of air pollutants in various mammals*. *Toxicology and Applied Pharmacology*, 238(1): p. 47-55.

17. Nichols J., et al., 2007. *Use of in vitro absorption, distribution, metabolism, and excretion (ADME) data in bioaccumulation assessments for fish*. *Human and Ecological Risk Assessment: An International Journal*, 13(6): p. 1164 - 1191.

18. McLachlan M.S., et al., 2011. *Bioaccumulation of organic contaminants in humans: A multimedia perspective and the importance of biotransformation*. *Environmental Science & Technology*, 45: p. 197-202.

19. Lipscomb J.C. and Poet T.S., 2008. *In vitro measurements of metabolism for application in pharmacokinetic modeling*. *Pharmacology & Therapeutics*, 118(1): p. 82-103.

20. van der Linde A., et al., 2001. *Estimating biotransformation rate constants of organic chemicals from modeled and measured elimination rates*. *Chemosphere*, 44(3): p. 423-435.

21. Arnot J.A., et al., 2014. *Estimating screening-level organic chemical half-lives in humans*. *Environmental Science & Technology*, 48(1): p. 723-730.

22. de Wolf W., et al., 2007. *Animal use replacement, reduction, and refinement: Development of an integrated testing strategy for bioconcentration of chemicals in fish*. *Integrated Environmental Assessment and Management*, 3(1): p. 3-17.

23. Houston J.B., 1994. *Utility of in vitro drug metabolism data in predicting in vivo metabolic clearance*. *Biochemical Pharmacology*, 47(9): p. 1469-1479.

24. Nelson D.L. and Cox M.M., 2005. *Lehninger Principles of biochemistry*. 4th ed. W. H. Freeman and Company: New York, NY, USA. 1119.

25. Testa B., et al., 2000. *The influence of lipophilicity on the pharmacokinetic behavior of drugs: Concepts and examples*. Perspectives in Drug Discovery and Design, 19(1): p. 179-211.
26. Hansch C., et al., 2004. *QSAR of cytochrome P450*. Drug Metabolism Reviews, 36(1): p. 105-156.
27. Lewis D.F.V., 2003. *Quantitative structure–activity relationships (QSARs) within the cytochrome P450 system: QSARs describing substrate binding, inhibition and induction of P450s*. Inflammopharmacology, 11(1): p. 43-73.
28. Long A. and Walker J.D., 2003. *Quantitative structure-activity relationships for predicting metabolism and modeling cytochrome P450 enzyme activities*. Environmental Toxicology and Chemistry, 22(8): p. 1894-1899.
29. Lewis D.F.V. and Dickins M., 2002. *Factors influencing rates and clearance in P450-mediated reactions: QSARs for substrates of the xenobiotic-metabolizing hepatic microsomal P450s*. Toxicology, 170(1-2): p. 45-53.
30. Lewis D.F.V., et al., 2004. *Compound lipophilicity for substrate binding to human P450s in drug metabolism*. Drug Discovery Today, 9(12): p. 530-537.
31. Cnubben N.H.P., et al., 1994. *Molecular orbital-based quantitative structure-activity relationship for the cytochrome P450-catalyzed 4-hydroxylation of halogenated anilines*. Chemical Research in Toxicology, 7(5): p. 590-598.
32. Strolin Benedetti M., et al., 2006. *Involvement of enzymes other than CYPs in the oxidative metabolism of xenobiotics*. Expert Opinion on Drug Metabolism & Toxicology, 2(6): p. 895-921.
33. Chang C., et al., 2009. *The development and validation of a computational model to predict rat liver microsomal clearance*. Journal of Pharmaceutical Sciences, 98(8): p. 2857-2867.
34. Ekins S. and Obach R.S., 2000. *Three-dimensional quantitative structure activity relationship computational approaches for prediction of human in vitro intrinsic clearance*. Journal of Pharmacology and Experimental Therapeutics, 295(2): p. 463-473.
35. Li H., et al., 2009. *First-principle, structure-based prediction of hepatic metabolic clearance values in human*. European Journal of Medicinal Chemistry, 44(4): p. 1600-1606.
36. Paixão P., et al., 2010. *Prediction of the in vitro intrinsic clearance determined in suspensions of human hepatocytes by using artificial neural networks*. European Journal of Pharmaceutical Sciences, 39(5): p. 310-321.

37. Madden J.C. and Cronin M.T.D., 2006. *Structure-based methods for the prediction of drug metabolism*. Expert Opinion on Drug Metabolism & Toxicology, 2(4): p. 545-557.
38. Ginsberg G., et al., 2004. *Incorporating pharmacokinetic differences between children and adults in assessing children's risks to environmental toxicants*. Toxicology and Applied Pharmacology, 198(2): p. 164-183.
39. Kaiser J.P., et al., 1996. *Microbial metabolism of pyridine, quinoline, acridine, and their derivatives under aerobic and anaerobic conditions*. Microbiological Reviews, 60(3): p. 483-98.
40. Safe S.H., 1994. *Polychlorinated biphenyls (PCBs): Environmental impact, biochemical and toxic responses, and implications for risk assessment*. Crit. Rev. Toxicol., 24(2): p. 87-149.
41. Snyder R. and Hedli C.C., 1996. *An overview of benzene metabolism*. Environmental Health Perspectives, 104(Suppl 6): p. 1165-1171.
42. Snedeker S.M., 2001. *Pesticides and breast cancer risk: A review of DDT, DDE, and dieldrin*. Environmental Health Perspectives, 109 (Suppl 1): p. 35-47.
43. Xue W. and Warshawsky D., 2005. *Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: A review*. Toxicology and Applied Pharmacology, 206(1): p. 73-93.
44. Seo J.S., et al., 2009. *Bacterial degradation of aromatic compounds*. International Journal of Environmental Research and Public Health, 6(1): p. 278-309.
45. Walther B., et al., 2008. *Lipophilicity of metabolites and its role in biotransformation*, in *Lipophilicity in Drug Action and Toxicology*. Wiley-VCH Verlag GmbH. p. 253-261.
46. Fears R., 1985. *Lipophilic xenobiotic conjugates: The pharmacological and toxicological consequences of the participation of drugs and other foreign compounds as substrates in lipid biosynthesis*. Progress in Lipid Research, 24(3): p. 177-195.
47. Giroud Y., et al., 1998. *Intrinsic and intramolecular lipophilicity effects in O-glucuronides*. Helvetica Chimica Acta, 81(2): p. 330-341.
48. Klaassen C.D., 2008. *Casarett and Doull's toxicology: The basic science of poisons - 7th ed*. McGraw-Hills.
49. Guengerich F.P., 2001. *Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity*. Chemical Research in Toxicology, 14(6): p. 611-650.

50. Brown C.M., et al., 2008. *Cytochromes P450: A structure-based summary of biotransformations using representative substrates*. Drug Metabolism Reviews, 40(1): p. 1-100.
51. Petrauskas A. and Kolovanov E., 2000. *ACD/Log P method description*. Perspectives in Drug Discovery and Design, 19(1): p. 99-116.
52. Machatha S.G. and Yalkowsky S.H., 2005. *Comparison of the octanol/water partition coefficients calculated by ClogP®, ACDlogP and KowWin® to experimentally determined values*. International Journal of Pharmaceutics, 294(1-2): p. 185-192.
53. Lowry R. *Concepts and Applications of Inferential Statistics*, Available from <http://faculty.vassar.edu/lowry/webtext.html> (accessed February 2012). 2012 February 2012]; Available from: <http://faculty.vassar.edu/lowry/webtext.html>.
54. Kollock R., et al., 2008. *Oxidation of alcohols and reduction of aldehydes derived from methyl- and dimethylpyrenes by cDNA-expressed human alcohol dehydrogenases*. Toxicology, 245(1-2): p. 65-75.
55. Caron G., et al., 1997. *Lipophilicity behavior of model and medicinal compounds containing a sulfide, sulfoxide, or sulfone moiety*. Helvetica Chimica Acta, 80(2): p. 449-462.
56. Caron G., et al., 1999. *Structure-property relationships in the basicity and lipophilicity of arylalkylamine oxides*. Helvetica Chimica Acta, 82(10): p. 1630-1639.
57. Scheer M., et al., 2011. *BRENDA, the enzyme information system in 2011*. Nucleic Acids Research, 39(suppl 1): p. D670-D676.
58. Krueger S.K. and Williams D.E., 2005. *Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism*. Pharmacology & Therapeutics, 106(3): p. 357-387.
59. Hansch C. and Zhang L., 1993. *Quantitative structure-activity relationships of cytochrome P-450*. Drug Metabolism Reviews, 25(1-2): p. 1-48.
60. Lewis D.F.V., 1999. *Frontier orbitals in chemical and biological activity: Quantitative relationships and mechanistic implication*. Drug Metabolism Reviews, 31(3): p. 755-816.
61. Weininger D., 1988. *SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules*. Journal of Chemical Information and Computer Sciences, 28(1): p. 31-36.
62. Madden J.C., et al., 2009. *Pharmaceuticals in the environment: Good practice in predicting acute ecotoxicological effects*. Toxicology Letters, 185(2): p. 85-101.



63. Zvinavashe E., et al., 2009. *On the number of EINECS compounds that can be covered by (Q)SAR models for acute toxicity*. Toxicology Letters, 184(1): p. 67-72.
64. Organisation for Economic Co-operation and Development, 2006. *Report on the Regulatory Uses and Applications in OECD Member Countries of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models in the Assessment of New and Existing Chemicals*. Paris, France.
65. Hendriks A.J., et al., 2005. *Critical body residues linked to octanol-water partitioning, organism composition, and LC50 QSARs: Meta-analysis and model*. Environmental Science & Technology, 39(9): p. 3226-3236.
66. Lewis D.F.V., 2000. *Structural characteristics of human P450s involved in drug metabolism: QSARs and lipophilicity profiles*. Toxicology, 144(1-3): p. 197-203.
67. Cronin M.T.D. and Schultz T.W., 2003. *Pitfalls in QSAR*. Journal of Molecular Structure (Theochem), 622(1-2): p. 39-51.
68. Deetz J.S., et al., 1984. *Human liver alcohol dehydrogenase isozymes: Reduction of aldehydes and ketones*. Biochemistry, 23(26): p. 6822-6828.
69. Holmes R., 1993. *Alcohol dehydrogenases: Gene multiplicity and differential functions of five classes of isozymes*. Drug and Alcohol Review, 12(1): p. 99-110.
70. Klyosov A.A., 1996. *Kinetics and specificity of human liver aldehyde dehydrogenases toward aliphatic, aromatic, and fused polycyclic aldehydes*. Biochemistry, 35(14): p. 4457-4467.
71. Lewis D.F.V. and Dickins M., 2003. *Baseline lipophilicity relationships in human cytochromes P450 associated with drug metabolism*. Drug Metabolism Reviews, 35(1): p. 1-18.
72. deBruyn A.M.H. and Gobas F.A.P.C., 2007. *The sorptive capacity of animal protein*. Environmental Toxicology and Chemistry, 26(9): p. 1803-1808.
73. Schwarzenbach R.P., et al., 2002. *Sorption II: Partitioning to living media - Bioaccumulation and baseline toxicity*, in *Environmental Organic Chemistry*. Wiley-Interscience: New York. p. 331-386.
74. Ziegler D.M., 1990. *Flavin-containing monooxygenases: Enzymes adapted for multisubstrate specificity*. Trends in Pharmacological Sciences, 11(8): p. 321-324.
75. Bu H.Z., 2006. *A literature review of enzyme kinetic parameters for CYP3A4-mediated metabolic reactions of 113 drugs in human liver microsomes: Structure- kinetics relationship assessment*. Current Drug Metabolism, 7(3): p. 231-249.

76. Weisbrod A.V., et al., 2009. *The state of in vitro science for use in bioaccumulation assessments for fish*. Environmental Toxicology and Chemistry, 28(1): p. 86-96.
77. Nichols J.W., et al., 2006. *In vitro-in vivo extrapolation of quantitative hepatic biotransformation data for fish: I. A review of methods, and strategies for incorporating intrinsic clearance estimates into chemical kinetic models*. Aquatic Toxicology, 78(1): p. 74-90.
78. Garrett R. and Grisham C.M., 2010. *Biochemistry*. 4th ed. Brooks/Cole, Cengage Learning: Boston, MA, USA. 1184.
79. Pirovano A., et al., 2012. *Compound lipophilicity as a descriptor to predict binding affinity (1/Km) in mammals*. Environmental Science & Technology, 46(9): p. 5168-5174.
80. Kikonyogo A. and Pietruszko R., 1996. *Aldehyde dehydrogenase from adult human brain that dehydrogenates gamma-aminobutyraldehyde: Purification, characterization, cloning and distribution*. Biochemical Journal, 316(part 1): p. 317-324.
81. Nagata T., et al., 1990. *Substrate specificities of rabbit lung and porcine liver flavin-containing monooxygenases: Differences due to substrate size*. Chemical Research in Toxicology, 3(4): p. 372-376.
82. Poulsen L.L. and Ziegler D.M., 1995. *Multisubstrate flavin-containing monooxygenases: Applications of mechanism to specificity*. Chemico-Biological Interactions, 96(1): p. 57-73.
83. Poulsen L.L., et al., 1979. *S-oxygenation of N-substituted thioureas catalyzed by the pig liver microsomal FAD-containing monooxygenase*. Archives of Biochemistry and Biophysics, 198(1): p. 78-88.
84. Smyser B.P. and Hodgson E., 1985. *Metabolism of phosphorus-containing compounds by pig liver microsomal FAD-containing monooxygenase*. Biochemical Pharmacology, 34(8): p. 1145-1150.
85. Taylor K.L. and Ziegler D.M., 1987. *Studies on substrate specificity of the hog liver flavin-containing monooxygenase: Anionic organic sulfur compounds*. Biochemical Pharmacology, 36(1): p. 141-146.
86. Ziegler D.M., 1988. *Flavin-containing monooxygenases: Catalytic mechanism and substrate specificities*. Drug Metabolism Reviews, 19(1): p. 1-32.
87. US EPA, *Estimation Programs Interface Suite™ for Microsoft® Windows*, v 4.1, Agency U.S.E.P., Editor.: Washington, DC, USA, <http://www.epa.gov/opptintr/exposure/pubs/episuite.htm>.
88. Black S.D. and Coon M.J., 1987. *P-450 cytochromes: Structure and function*, in *Advances in Enzymology and Related Areas of Molecular Biology*,

Volume 60, Alton M. (Ed.). John Wiley and Sons, Inc.: Hoboken, NJ, USA. p. 35-88.

89. Lewis D.F.V., 2000. *On the recognition of mammalian microsomal cytochrome P450 substrates and their characteristics: Towards the prediction of human P450 substrate specificity and metabolism*. *Biochemical Pharmacology*, 60(3): p. 293-306.

90. Dearden J.C., 1990. *Physico-chemical descriptors*, in *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*, Karcher W. and Devillers J. (Eds.). Kluwer Academic Publishers: Dordrecht, The Netherlands. p. 25-60.

91. Sushko I., et al., 2011. *Online chemical modeling environment (OCHEM): Web platform for data storage, model development and publishing of chemical information*. *Journal of Computer-Aided Molecular Design*, 25(6): p. 533-554.

92. Stewart J.J.P., *MOPAC2009 Version 11.366W*. 2008, Stewart Computational Chemistry: Colorado Springs, CO, USA.

93. Pedretti A., et al., 2004. *VEGA - An open platform to develop chemo-bio-informatics applications, using plug-in architecture and script programming*. *Journal of Computer-Aided Molecular Design*, 18(3): p. 167-173.

94. Lewis D.F.V., 1997. *Quantitative structure-activity relationships in substrates, inducers, and inhibitors of cytochrome P4501 (CYP1)*. *Drug Metabolism Reviews*, 29(3): p. 589-650.

95. R Core Team, *R: A Language and Environment for Statistical Computing*. 2012, R Foundation for Statistical Computing: Vienna, Austria.

96. McLeod A.I. and Xu C. *bestglm: Best subset GLM*. 2011; Available from: <http://cran.r-project.org/web/packages/bestglm/index.html>.

97. Dearden J.C., et al., 2009. *How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR)*. *SAR and QSAR in Environmental Research*, 20(3): p. 241-266.

98. Fox J., et al. *car: Companion to Applied Regression*. 2012; Available from: <http://cran.r-project.org/web/packages/car/index.html>.

99. Zuur A.F., et al., 2007. *Analysing Ecological Data*. Springer: New York, NY, USA. 698.

100. Zuur A.F., et al., 2009. *Mixed Effects Models and Extensions in Ecology with R*. Springer: New York, NY, USA. 574.

101. Hall M., et al., 2009. *The WEKA data mining software: An update*. *SIGKDD Explorations*, 11(1): p. 10-18.

102. Waller C.L., et al., 1996. *Modeling the cytochrome P450-mediated metabolism of chlorinated volatile organic compounds*. Drug Metabolism and Disposition, 24(2): p. 203-210.
103. Todeschini R., et al., 2009. *Chemometrics in QSAR*, in *Comprehensive Chemometrics, volume 4*, Brown S., Tauler R., and Walczak R. (Eds.). Elsevier: Oxford, UK. p. 129-172.
104. Zvinavashe E., et al., 2008. *Promises and pitfalls of Quantitative Structure–Activity Relationship approaches for predicting metabolism and toxicity*. Chemical Research in Toxicology, 21(12): p. 2229-2236.
105. Vasiliou V., et al., 2000. *Role of aldehyde dehydrogenases in endogenous and xenobiotic metabolism*. Chemico-Biological Interactions, 129(1-2): p. 1-19.
106. Shimada T., et al., 2013. *Binding of diverse environmental chemicals with human cytochromes P450 2A13, 2A6, and 1B1 and enzyme inhibition*. Chemical Research in Toxicology, 26: p. 517-528.
107. Guengerich F.P., et al., 1995. *Interpretations of cytochrome P450 mechanisms from kinetic studies*. Biochimie, 77(7–8): p. 573-580.
108. Damborsky J. and Wayne S.T., 1997. *Comparison of the QSAR models for toxicity and biodegradability of anilines and phenols*. Chemosphere, 34(2): p. 429-446.
109. Karelson M. and Lobanov V.S., 1996. *Quantum-chemical descriptors in QSAR/QSPR studies*. Chemical Reviews, 96(3): p. 1027-1044.
110. Enoch S.J., 2010. *The use of quantum mechanics derived descriptors in computational toxicology*, in *Recent Advances in QSAR Studies, Methods and Applications, Volume 8*, Puzyn T., Leszczynski J., and Cronin M.T. (Eds.). Springer: Dordrecht, The Netherlands. p. 13-28.
111. Garcia-Viloca M., et al., 2004. *How enzymes work: Analysis by modern rate theory and computer simulations*. Science, 303(5655): p. 186-195.
112. Farrés J., et al., 1995. *Investigation of the active site cysteine residue of rat liver mitochondrial aldehyde dehydrogenase by site-directed mutagenesis*. Biochemistry, 34(8): p. 2592-2598.
113. Wang Y., et al., 2014. *Investigation on the relationship between bioconcentration factor and distribution coefficient based on class-based compounds: The factors that affect bioconcentration*. Environmental Toxicology and Pharmacology, 38(2): p. 388-396.
114. Blaauboer B.J., 2002. *The applicability of in vitro-derived data in hazard identification and characterisation of chemicals*. Environmental Toxicology and Pharmacology, 11(3–4): p. 213-225.

115. Cherkasov A., et al., 2013. *QSAR modeling: Where have you been? Where are you going to?* Journal of Medicinal Chemistry, 57(12): p. 4977-5010.
116. Balaz S., 2009. *Modeling kinetics of subcellular disposition of chemicals.* Chemical Reviews, 109(5): p. 1793-1899.
117. Pirovano A., et al., 2014. *Mechanistically-based QSARs to describe metabolic constants in mammals.* ATLA, 42(1): p. 59–69.
118. Consonni V. and Todeschini R., 2010. *Molecular Descriptors*, in *Recent Advances in QSAR Studies: Methods and Applications*, Puzyn T., Leszczynski J., and Cronin M.T.D. (Eds.). Springer. p. 29-102.
119. Stewart J.P., 1990. *MOPAC: A semiempirical molecular orbital program.* Journal of Computer-Aided Molecular Design, 4(1): p. 1-103.
120. Hall L.H. and Kier L.B., 1995. *Electrotopological state indices for atom types: A novel combination of electronic, topological, and valence state information.* Journal of Chemical Information and Computer Sciences, 35(6): p. 1039-1045.
121. Tetko I.V. and Tanchuk V.Y., 2002. *Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program.* Journal of Chemical Information and Computer Sciences, 42(5): p. 1136-1145.
122. Steinbeck C., et al., 2006. *Recent developments of the Chemistry Development Kit (CDK) - An open-source Java library for chemo- and bioinformatics.* Current Pharmaceutical Design, 12(17): p. 2111-2120.
123. Mauri A., et al., 2006. *DRAGON software: an easy approach to molecular descriptor calculations.* MATCH Communications in Mathematical and in Computer Chemistry, 56: p. 237-248.
124. Papa E., et al., 2014. *Metabolic biotransformation half-lives in fish: QSAR modeling and consensus analysis.* Science of The Total Environment, 470-471: p. 1040-1046.
125. Frank E., et al., 2010. *Weka-A Machine Learning Workbench for Data Mining*, in *Data Mining and Knowledge Discovery Handbook*, Maimon O. and Rokach L. (Eds.). Springer US. p. 1269-1277.
126. Leardi R., 2001. *Genetic algorithms in chemometrics and chemistry: A review.* Journal of Chemometrics, 15(7): p. 559-569.
127. Fox J., et al. *car: Companion to Applied Regression.* 2014; Available from: <http://cran.r-project.org/web/packages/car/index.html>.
128. Morrill J.A., et al., 2011. *Development of quantitative structure–activity relationships for explanatory modeling of fast reacting (meth)acrylate monomers bearing novel functionality.* J. Mol. Graphics Modell, 29(5): p. 763-772.

129. Eriksson L., et al., 2003. *Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs*. Environmental Health Perspectives, 111(10): p. 1361-1375.
130. Hodgson E., et al., 1995. *Pesticide-metabolizing enzymes*. Toxicology Letters, 82-83: p. 73-81.
131. Buters J.T.M., 2008. *Phase I Metabolism*, in *Toxicology and Risk Assessment: A Comprehensive Introduction* Greim H. and Snyder R. (Eds.). Wiley & Sons Ltd. p. 49-73.
132. Todeschini R. and Consonni V., 2009. *Molecular Descriptors for Chemoinformatics, 2nd edition*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.
133. Galvez J., et al., 1994. *Charge indexes. New topological descriptors*. Journal of Chemical Information and Computer Sciences, 34(3): p. 520-525.
134. Liu H. and Gramatica P., 2007. *QSAR study of selective ligands for the thyroid hormone receptor  $\beta$* . Bioorganic & Medicinal Chemistry, 15(15): p. 5251-5261.
135. Todeschini R. and Consonni V., 2003. *Descriptors from Molecular Geometry*, in *Handbook of Chemoinformatics*, Gasteiger J. (Ed.). Wiley-VCH Verlag GmbH: Weinheim, Germany. p. 1004-1033.
136. Strempel S., et al., 2013. *Using conditional inference trees and random forests to predict the bioaccumulation potential of organic chemicals*. Environmental Toxicology and Chemistry: p. 1187-1195.
137. Cowan-Ellsberry C.E., et al., 2008. *Approach for extrapolating in vitro metabolism data to refine bioconcentration factor estimates*. Chemosphere, 70(10): p. 1804-1817.
138. Nichols J.W., et al., 2013. *Toward improved models for predicting bioconcentration of well-metabolized compounds by rainbow trout using measured rates of in vitro intrinsic clearance*. Environmental Toxicology and Chemistry, 32(7): p. 1611-1622.
139. Han X., et al., 2007. *Determination of xenobiotic intrinsic clearance in freshly isolated hepatocytes from rainbow trout (*Oncorhynchus mykiss*) and rat and its application in bioaccumulation assessment*. Environmental Science & Technology, 41(9): p. 3269-3276.
140. Han X., et al., 2009. *Liver microsomes and S9 from rainbow trout (*Oncorhynchus mykiss*): Comparison of basal-level enzyme activities with rat and determination of xenobiotic intrinsic clearance in support of bioaccumulation assessment*. Environmental Toxicology and Chemistry, 28(3): p. 481-488.

141. Pirovano A., et al., 2015. *The utilisation of structural descriptors to predict metabolic constants of xenobiotics in mammals*. Environmental Toxicology and Pharmacology, 39(1): p. 247-258.
142. Coe K.J. and Koudriakova T., 2014. *Metabolic stability assessed by liver microsomes and hepatocytes*, in *Optimization in Drug Discovery*, Caldwell G.W. and Yan Z. (Eds.). Humana Press. p. 87-99.
143. Di L., et al., 2012. *Mechanistic insights from comparing intrinsic clearance values between human liver microsomes and hepatocytes to guide drug design*. European Journal of Medicinal Chemistry, 57(0): p. 441-448.
144. Ito K. and Houston J.B., 2004. *Comparison of the use of liver models for predicting drug clearance using in vitro kinetic data from hepatic microsomes and isolated hepatocytes*. Pharmaceutical Research, 21(5): p. 785-792.
145. Tonnelier A., et al., 2012. *Screening of chemicals for human bioaccumulative potential with a physiologically based toxicokinetic model*. Archives of Toxicology, 86(3): p. 393-403.
146. Sohlenius-Sternbeck A.-K., et al., 2010. *Evaluation of the human prediction of clearance from hepatocyte and microsome intrinsic clearance for 52 drug compounds*. Xenobiotica, 40(9): p. 637-649.
147. Gramatica P., et al., 2012. *QSAR modeling is not “push a button and find a correlation”: A case study of toxicity of (benzo-)triazoles on algae*. Molecular Informatics, 31(11-12): p. 817-835.
148. Fourches D., et al., 2010. *Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research*. Journal of Chemical Information and Modeling, 50(7): p. 1189-1204.
149. Hall L.H. and Kier L.B., 2000. *The E-state as the basis for molecular structure space definition and structure similarity*. Journal of Chemical Information and Computer Sciences, 40(3): p. 784-791.
150. Marvin 5, 2012, ChemAxon (<http://www.chemaxon.com>).
151. Copley S.D., 2000. *Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach*. Trends in Biochemical Sciences, 25(6): p. 261-265.
152. Ritter R., et al., 2011. *Intrinsic human elimination half-lives of polychlorinated biphenyls derived from the temporal evolution of cross-sectional biomonitoring data from the United Kingdom*. Environmental Health Perspectives, 119(2): p. 225-231.
153. Tang J., et al., 2006. *Metabolism of organophosphorus and carbamate pesticides*, in *Toxicology of Organophosphate & Carbamate Compounds*, Gupta R.C. (Ed.). Academic Press: Burlington. p. 127-143.

154. Barter Z.E., et al., 2007. *Scaling factors for the extrapolation of in vivo metabolic drug clearance from in vitro data: reaching a consensus on values of human micro-somal protein and hepatocellularity per gram of liver*. Current Drug Metabolism, 8: p. 33-45.
155. Davies B. and Morris T., 1993. *Physiological parameters in laboratory animals and humans*. Pharmaceutical Research, 10(7): p. 1093-1095.
156. Bale A.S., et al., 2013. *Correlating in vitro data to in vivo findings for risk assessment*. ALTEX, 31(1): p. 79-90.
157. De Buck S.S., et al., 2007. *The prediction of drug metabolism, tissue distribution, and bioavailability of 50 structurally diverse compounds in rat using mechanism-based absorption, distribution, and metabolism prediction tools*. Drug Metabolism and Disposition, 35(4): p. 649-659.
158. De Buck S.S., et al., 2007. *Prediction of human pharmacokinetics using physiologically based modeling: A retrospective analysis of 26 clinically tested drugs*. Drug Metabolism and Disposition, 35(10): p. 1766-1780.
159. Lobell M. and Sivarajah V., 2003. *In silico prediction of aqueous solubility, human plasma protein binding and volume of distribution of compounds from calculated  $pK_a$  and AlogP98 values*. Molecular Diversity, 7(1): p. 69-87.
160. Goodman L.S., et al., 2006. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*. 11th ed.: New York.
161. Naritomi Y., et al., 2001. *Prediction of human hepatic clearance from in vivo animal experiments and in vitro metabolic studies with liver microsomes from animals and humans*. Drug Metabolism and Disposition, 29(10): p. 1316-1324.
162. Riley R.J., et al., 2005. *A unified model for predicting human hepatic, metabolic clearance from in vitro intrinsic clearance data in hepatocytes and microsomes*. Drug Metabolism and Disposition, 33(9): p. 1304-1311.
163. Krause P., et al., 2005. *Comparison of different efficiency criteria for hydrological model assessment*. Advances in Geosciences, 5: p. 89-97.
164. Kirchmair J., et al., 2013. *How Do Metabolites Differ from Their Parent Molecules and How Are They Excreted?* Journal of Chemical Information and Modeling, 53(2): p. 354-367.
165. Kirchmair J., et al., 2015. *Predicting drug metabolism: experiment and/or computation?* Nature Reviews Drug Discovery.
166. Parkinson A. and Safe S., 1982. *Cytochrome P-450-mediated metabolism of biphenyl and the 4-halobiphenyls*. Biochemical Pharmacology, 31(10): p. 1849-1856.



167. Sundström G., et al., 1976. *The metabolism of chlorobiphenyls — A review*. Chemosphere, 5(5): p. 267-298.
168. Stadnicki S.S. and Allen J.R., 1979. *Toxicity of 2,2',5,5'-tetrachlorobiphenyl and its metabolites, 2,2',5,5'-tetrachlorobiphenyl-3,4-oxide and 2,2',5,5'-tetrachlorobiphenyl-4-ol to cultured cells in vitro*. Bulletin of Environmental Contamination and Toxicology, 23(1): p. 788-796.
169. James M.O., 2001. *Polychlorinated biphenyls: Metabolism and metabolites*, in *PCBs: Recent Advances in Environmental Toxicology and Health Effects*, Larry W. Robertson L.G.H. (Ed.). The University Press of Kentucky. p. 35-46.
170. Hovander L., et al., 2002. *Identification of hydroxylated PCB metabolites and other phenolic halogenated pollutants in human blood plasma*. Archives of Environmental Contamination and Toxicology, 42(1): p. 105-117.
171. Letcher R., et al., 2000. *Methyl Sulfone and Hydroxylated Metabolites of Polychlorinated Biphenyls*, in *The Handbook of Environmental Chemistry*, Hutzinger O. and Paasivirta J. (Eds.). Springer Berlin / Heidelberg. p. 315-359.
172. Fängström B., et al., 2002. *Hydroxylated PCB metabolites and PCBs in serum from pregnant Faroese women*. Environ Health Perspect, 110(9).
173. Koga N., et al., 1990. *Metabolism in vivo of 3,4,5,3', 4'-pentachlorobiphenyl and toxicological assessment of the metabolite in rats*. Journal of Pharmacobio-Dynamics, 13: p. 497-506.
174. Koop D., 1992. *Oxidative and reductive metabolism by cytochrome P450 2E1*. The FASEB Journal, 6(2): p. 724-730.
175. Berner T., et al., 2009. *Toxicological review of nitrobenzene (CAS No. 98-95-3)*.
176. Midorikawa K., et al., 2004. *Metabolic activation of carcinogenic ethylbenzene leads to oxidative DNA damage*. Chemico-Biological Interactions, 150(3): p. 271-281.
177. Pedersen R.T. and Hill E.M., 2000. *Biotransformation of the xenoestrogen 4-tert-octylphenol in hepatocytes of rainbow trout (Oncorhynchus mykiss)*. Xenobiotica, 30(9): p. 867-879.
178. Yan Z., et al., 2005. *Bioactivation of 4-methylphenol (p-cresol) via cytochrome p450-mediated aromatic oxidation in human liver microsomes*. Drug Metabolism and Disposition, 33(12): p. 1867-1876.
179. Zerilli A., et al., 1997. *Both cytochromes P450 2E1 and 3A are involved in the o-hydroxylation of p-nitrophenol, a catalytic activity known to be specific for P450 2E1*. Chemical Research in Toxicology, 10(10): p. 1205-1212.

180. Daly J.W., et al., 1968. *Hydroxylation of alkyl and halogen substituted anilines and acetanilides by microsomal hydroxylases*. *Biochemical Pharmacology*, 17(1): p. 31-36.
181. LaVoie E.J., et al., 1983b. *On the metabolism of quinoline and isoquinoline: possible molecular basis for differences in biological activities*. *Carcinogenesis*, 4(9): p. 1169-1173.
182. LaVoie E.J., et al., 1983a. *Identification of the metabolites of benzo[f]quinoline and benzo[h]-quinoline formed by rat liver homogenate*. *Carcinogenesis*, 4(9): p. 1133-1138.
183. Murphy S.E., et al., 1992. *Rat liver metabolism of benzo[b]naphtho[2,1-d]thiophene*. *Chemical Research in Toxicology*, 5(4): p. 491-495.
184. Perin F., et al., 1981. *Heterocyclic polycyclic aromatic hydrocarbon carcinogenesis: 7H-dibenzo[c,g]carbazole metabolism by microsomal enzymes from mouse and rat liver*. *Chemico-Biological Interactions*, 35(3): p. 267-284.
185. Wilke T.J., et al., 1989. *Oxidative metabolism of <sup>14</sup>C-pyridine by human and rat tissue subcellular fractions*. *Xenobiotica*, 19(9): p. 1013-1022.
186. Van den Berg M., et al., 1994. *The toxicokinetics and metabolism of polychlorinated dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs) and their relevance for toxicity*. *Critical Reviews in Toxicology*, 24(1): p. 1-74.
187. Inouye K., et al., 2002. *Metabolism of polychlorinated dibenzo-p-dioxins (PCDDs) by human cytochrome P450-dependent monooxygenase systems*. *Journal of Agricultural and Food Chemistry*, 50(19): p. 5496-5502.
188. Petroske E., et al., 1997. *Identification of NIH-shifted metabolites of 1,3,7,8-tetrachlorodibenzo-p-dioxin in the rat by NMR comparison with synthesized isomers*. *Chemosphere*, 34(5-7): p. 1549-1555.
189. Hamers T., et al., 2008. *Biotransformation of brominated flame retardants into potentially endocrine-disrupting metabolites, with special attention to 2,2',4,4'-tetrabromodiphenyl ether (BDE-47)*. *Molecular Nutrition & Food Research*, 52(2): p. 284-298.
190. Stapleton H., et al., 2009. *Metabolism of polybrominated diphenyl ethers (PBDEs) by human hepatocytes in vitro*. *Environmental Health Perspectives*, 117(2): p. 197-202.
191. Hakk H. and Letcher R.J., 2003. *Metabolism in the toxicokinetics and fate of brominated flame retardants-a review*. *Environment International*, 29(6): p. 801-828.
192. Couri D. and Milks M., 1982. *Toxicity and metabolism of the neurotoxic hexacarbons n-hexane, 2-hexanone, and 2,5-hexanedione*. *Annual Review of Pharmacology and Toxicology*, 22: p. 145-66.

193. Das M.L., et al., 1968. *On the fatty acid and hydrocarbon hydroxylation in rat liver microsomes*. European Journal of Biochemistry, 4(4): p. 519-523.
194. Cravedi J.P., et al., 1989. *Hydroxylation of pristane by isolated hepatocytes of rainbow trout: A comparison with in vivo metabolism and biotransformation by liver microsomes*. Marine Environmental Research, 28(1-4): p. 15-18.
195. Snyder R., 1992. *Ethel Browning's Toxicity and Metabolism of Industrial Solvents*. 2nd ed, ed. Snyder R. Vol. Volume 3: Alcohols and Esters. Elsevier: New York, NY.
196. Nakajima T., et al., 1997. *Toluene metabolism by cDNA-Expressed human hepatic cytochrome P450*. Biochemical Pharmacology, 53(3): p. 271-277.
197. Lewis D.F.V., et al., 2003. *A quantitative structure-activity relationship analysis on a series of alkyl benzenes metabolized by human cytochrome P450 2E1*. Journal of Biochemical and Molecular Toxicology, 17(1): p. 47-52.
198. Meldahl A.C., et al., 1996. *Metabolism of several <sup>14</sup>C-nonylphenol isomers by rainbow trout (Oncorhynchus mykiss): In vivo and in vitro microsomal metabolites*. Xenobiotica, 26(11): p. 1167-1180.
199. Thibaut R., et al., 1998. *Characterization of biliary metabolites of 4-n-nonylphenol in rainbow trout (Oncorhynchus mykiss)*. Xenobiotica, 28(8): p. 745-757.
200. International Programme on Chemical Safety (IPCS). *IPCS, Poisons Information Monograph 095: Camphor*. 1989; Available from: <http://www.inchem.org/documents/pims/pharm/camphor.htm#SectionTitle:1.5%20Brand%20names,%20Trade%20na>.
201. Chen C. and Lin C.C., 1968. *Mechanism of aliphatic hydroxylation tetralin hydroperoxide as an intermediate in the hydroxylation of tetralin in rat-liver homogenate*. Biochimica et Biophysica Acta, General Subjects, 170(2): p. 366-374.
202. Fang J., et al., 1999. *Metabolism of risperidone to 9-hydroxyrisperidone by human cytochromes P450 2D6 and 3A4*. Naunyn-Schmiedeberg's Archives of Pharmacology, 359(2): p. 147-151.
203. Zhou S.-F., 2009. *Polymorphism of human cytochrome P450 2D6 and its clinical significance: Part I*. Clinical Pharmacokinetics, 48(11): p. 689-723.
204. Broly F., et al., 1990. *Mexiletine metabolism in vitro by human liver*. Drug Metabolism and Disposition, 18(3): p. 362-368.
205. Tang B.K., et al., 1980. *Species differences of amobarbital metabolism: dihydroxyamobarbital formation*. Canadian Journal of Physiology and Pharmacology, 58(10): p. 1167-1169.

206. Shimada T. and Fujii-Kuriyama Y., 2004. *Metabolic activation of polycyclic aromatic hydrocarbons to carcinogens by cytochromes P450 1A1 and 1B1*. Cancer Science, 95(1): p. 1-6.
207. Lau S.S. and Zannoni V.G., 1981. *Bromobenzene epoxidation leading to binding on macromolecular protein sites*. Journal of Pharmacology and Experimental Therapeutics, 219(2): p. 563-572.
208. Härkönen H., 1978. *Styrene, its experimental and clinical toxicology. A review*. Scandinavian Journal of Work, Environment & Health, 4 Suppl 2: p. 104-13.
209. Vaz A.D.N., et al., 1998. *Epoxidation of olefins by cytochrome P450: Evidence from site-specific mutagenesis for hydroperoxo-iron as an electrophilic oxidant*. Proceedings of the National Academy of Sciences, 95(7): p. 3555-3560.
210. Guengerich F.P., 2003. *Cytochrome P450 oxidations in the generation of reactive electrophiles: epoxidation and related reactions*. Archives of Biochemistry and Biophysics, 409(1): p. 59-71.
211. Born S.L., et al., 2002. *Identification of the cytochromes P450 that catalyze coumarin 3,4-epoxidation and 3-hydroxylation*. Drug Metabolism and Disposition, 30(5): p. 483-487.
212. Groves J.T., et al., 1986. *Hydrogen-deuterium exchange during propylene epoxidation by cytochrome P-450*. Journal of the American Chemical Society, 108(13): p. 3837-3838.
213. Ortiz de Montellano P.R., et al., 1983. *Stereochemistry of cytochrome P-450-catalyzed epoxidation and prosthetic heme alkylation*. Journal of Biological Chemistry, 258(7): p. 4208-13.
214. Gervasi P.G. and Longo V., 1990. *Metabolism and mutagenicity of isoprene*. Environmental Health Perspectives, 86: p. 85-87.
215. Woggon W.-D., 1997. *Cytochrome P450: Significance, Reaction Mechanisms and Active Site Analogues*, in *Topics in Current Chemistry*, Schmidtchen F. (Ed.). Springer Berlin / Heidelberg. p. 39-96.
216. Smith B.J., et al., 1990. *Comparison of the disposition and in vitro metabolism of 4-vinylcyclohexene in the female mouse and rat*. Toxicology and Applied Pharmacology, 105(3): p. 364-371.
217. Wolff T., et al., 1979. *Aldrin epoxidation, a highly sensitive indicator specific for cytochrome P-450-dependent mono-oxygenase activities*. Drug Metabolism and Disposition, 7(5): p. 301-305.
218. Sumner S.C.J., et al., 1999. *Role of cytochrome P450 2E1 in the metabolism of acrylamide and acrylonitrile in mice*. Chemical Research in Toxicology, 12(11): p. 1110-1116.

219. Guengerich F.P., 2007. *Cytochrome P450 and chemical toxicology*. Chemical Research in Toxicology, 21(1): p. 70-83.
220. Roberts-Thomson S.J., et al., 1995. *Metabolism of polycyclic aza-aromatic carcinogens catalyzed by four expressed human cytochromes P450*. Cancer Research, 55(5): p. 1052-1059.
221. Leibman K.C. and Ortiz E., 1968. *Oxidation of indene in liver microsomes*. Molecular Pharmacology, 4(3): p. 201-207.
222. Nordqvist M., et al., 1981. *Metabolism of chrysene and phenanthrene to bay-region diol epoxides by rat liver enzymes*. Molecular Pharmacology, 19(1): p. 168-178.
223. International Agency for Research on Cancer (IARC). *Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man*. 1983 [cited (Multivolume work); p. V32 229].
224. Jacob J., et al., 1982. *The metabolism of pyrene by rat liver microsomes and the influence of various mono-oxygenase inducers*. Xenobiotica, 12(1): p. 45-53.
225. Yang S.K., 1988. *Stereoselectivity of cytochrome P-450 isozymes and epoxide hydrolase in the metabolism of polycyclic aromatic hydrocarbons*. Biochemical Pharmacology, 37(1): p. 61-70.
226. MacNicoll A.D., et al., 1979. *The formation of dihydrodiols in the chemical or enzymic oxidation of dibenz[a,c]anthracene, dibenz[a,h]anthracene and chrysene*. Chemico-Biological Interactions, 27(2-3): p. 365-379.
227. Akhtar M.N., et al., 1979. *Anthracene 1,2-oxide: Synthesis and role in the metabolism of anthracene by mammals*. Journal of the Chemical Society, Perkin Transactions 1.
228. International Programme on Chemical Safety (IPCS). *IPCS, Environmental Health Criteria 229: Selected Nitro- and Nitro-Oxy-Polycyclic Aromatic Hydrocarbons*. 2003; Available from: <http://www.inchem.org/pages/ehc.html>.
229. Chou M.W., et al., 1986. *Metabolism of l-nitrobenzo[a]pyrene by rat liver microsomes to potent mutagenic metabolites*. Carcinogenesis, 7(11): p. 1837-1844.
230. Fu P.P., et al., 1985. *Metabolism of 9-nitroanthracene by rat liver microsomes: identification and mutagenicity of metabolites*. Carcinogenesis, 6(5): p. 753-757.
231. Takeshita M., et al., 1995. *Regio- and stereo-selective oxidation of phenylbutane by rat liver*. Research communications in molecular pathology and pharmacology, 89(3): p. 351-356.

232. Talaat R.E. and Nelson W.L., 1988. *Regioisomeric aromatic dihydroxylation of propranolol. Synthesis and identification of 4,6- and 4,8-dihydroxypropranolol as metabolites in the rat and in man.* Drug Metabolism and Disposition, 16(2): p. 212-216.
233. US Environmental Protection Agency (USEPA), 2004. *Interim Reregistration Eligibility Decision for Carbaryl.*
234. Fukami J.-i., et al., 1967. *Metabolism of rotenone in vitro by tissue homogenates from mammals and insects.* Science, 155(3763): p. 713-716.
235. Montesissa C., et al., 1995. *In vitro comparison of aldicarb oxidation in various food-producing animal species.* Veterinary and Human Toxicology, 37(4): p. 333-336.
236. Usmani K.A., et al., 2004a. *In vitro sulfoxidation of thioether compounds by human cytochrome P450 and flavin-containing monooxygenase isoforms with particular reference to the CYP2C family.* Drug Metabolism and Disposition, 32(3): p. 333-339.
237. Aizawa H., 1982. *Metabolic Maps of Pesticides.* Academic Press: New York, NY.
238. Tomlin C., 2002. *Ethiofencarb (29973-13-5) in "The e-Pesticide Manual", version 2.2, 12th edition.* British Crop. Production Council: Hampshire, UK.
239. Hayes W.J. and Laws E.R., 1991. *Handbook of Pesticide Toxicology. Volume 3: Classes of Pesticides.* Academic Press: New York, NY.
240. Menzie C.M., 1978. *Metabolism of Pesticides, Update II.* Washington, DC.
241. World Health Organization (WHO) and Food and Agriculture Organization of the United Nations (FAO). *Joint FAO/WHO Meeting on Pesticide Residues. Evaluation for Clethodim (99129-21-2).* 1994; Available from: <http://www.inchem.org/pages/jmpr.html>.
242. Leoni C., et al., 2008. *The participation of human hepatic P450 isoforms, flavin-containing monooxygenases and aldehyde oxidase in the biotransformation of the insecticide fenthion.* Toxicology and Applied Pharmacology, 233(2): p. 343-352.
243. Kulkarni A.P. and Hodgson E., 1980. *Metabolism of insecticides by mixed function oxidase systems.* Pharmacology & Therapeutics, 8(2): p. 379-475.
244. Goodwin B.L., 1976. *Handbook of Intermediary Metabolism of Aromatic Compounds.* Wiley: New York, NY.
245. Bingham E., et al., 2001. *Patty's Toxicology.* 5th ed. Vol. Volumes 1-9. John Wiley & Sons: New York, NY.

246. Spencer E.Y., 1982. *Guide to the Chemicals Used in Crop Protection*. 7th ed. Publication 1093. Research Institute, Agriculture Canada, Ottawa, Canada: Information Canada.
247. Lang D., et al., 1996. *In vitro metabolism of atrazine, terbutylazine, ametryne, and terbutryne in rats, pigs, and humans*. Drug Metabolism and Disposition, 24(8): p. 859-865.
248. Hammons G.J., et al., 1997. *Metabolism of carcinogenic heterocyclic and aromatic amines by recombinant human cytochrome P450 enzymes*. Carcinogenesis, 18(4): p. 851-854.
249. Shimada T., et al., 1996. *Activation of chemically diverse procarcinogens by human cytochrome P-450 1B1*. Cancer Research, 56(13): p. 2979-2984.
250. Duncan J.D. and Cho A.K., 1982. *N-oxidation of phentermine to N-hydroxyphentermine by a reconstituted cytochrome P-450 oxidase system from rabbit liver*. Molecular Pharmacology, 22(2): p. 235-238.
251. Department of Health and Human Services (DHHS) and Agency for Toxic Substances and Disease Registry (ATSDR). *Toxicological Profile for Methylenedianiline (PB/99/102568/AS)*. 1998; Available from: <http://www.atsdr.cdc.gov/ToxProfiles/tp.asp?id=1001&tid=210>.
252. International Agency for Research on Cancer (IARC). *Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man*. 1975; Available from: <http://monographs.iarc.fr/index.php>.
253. Snyder R., 1990. *Ethyl Browning's Toxicity and Metabolism of Industrial Solvents*, ed. Snyder R. Vol. Volume 2: Nitrogen and Phosphorus Solvents. Elsevier: Amsterdam-New York-Oxford.
254. Kim D. and Guengerich F.P., 2005. *Cytochrome P450 activation of arylamines and heterocyclic amines*. Annual Review of Pharmacology and Toxicology, 45(1): p. 27-49.
255. Hardman J.G., et al., 2001. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*. 10th ed. McGraw-Hill: New York, NY.
256. International Programme on Chemical Safety (IPCS). *IPCS, Environmental health criteria 102: 1-Propanol*. 1990; Available from: <http://www.inchem.org/documents/ehc/ehc/ehc102.htm>.
257. Organisation for Economic Co-operation and Development (OECD). *Screening Information Data Set (SIDS): Isobutanol*. 2004; Available from: <http://www.inchem.org/documents/sids/sids/78831.pdf>.
258. Hinson J.A. and Neal R.A., 1975. *An examination of octanol and octanal metabolism to octanoic acid by horse liver alcohol dehydrogenase*. Biochimica et Biophysica Acta, Enzymology, 384(1): p. 1-11.

259. Green T., et al., 2002. *The development of forestomach tumours in the mouse following exposure to 2-butoxyethanol by inhalation: Studies on the mode of action and relevance to humans*. Toxicology, 180(3): p. 257-273.
260. Fontaine F.R., et al., 2002. *Oxidative bioactivation of crotyl alcohol to the toxic endogenous aldehyde crotonaldehyde: Association of protein carbonylation with toxicity in mouse hepatocytes*. Chemical Research in Toxicology, 15(8): p. 1051-1058.
261. Keung W.-M., 1991. *Human liver alcohol dehydrogenases catalyze the oxidation of the intermediary alcohols of the shunt pathway of mevalonate metabolism*. Biochemical and Biophysical Research Communications, 174(2): p. 701-707.
262. International Programme on Chemical Safety (IPCS). *IPCS, Food Additives Series 52: Aliphatic branched-chain saturated and unsaturated alcohols, aldehydes, acids, and related esters. Annex n. 6: Flavouring agents with minimum assay values of less than 95%*. 2004; Available from: [http://libdoc.who.int/publications/2004/924166052X\\_annexes.pdf](http://libdoc.who.int/publications/2004/924166052X_annexes.pdf)  
<http://www.inchem.org/documents/jecfa/jecmono/v52je15.htm>.
263. Hartley D.P., et al., 1995. *The hepatocellular metabolism of 4-hydroxynonenal by alcohol dehydrogenase, aldehyde dehydrogenase, and glutathione S-transferase*. Archives of Biochemistry and Biophysics, 316(1): p. 197-205.
264. Dart R.C., 2004. *Medical Toxicology*. 3rd ed, ed. Dart R.C. Lippincott Williams & Wilkins: Philadelphia, PA.
265. Kalf G.F., et al., 1987. *Solvent toxicology: Recent advances in the toxicology of benzene, the glycol ethers, and carbon tetrachloride*. Annual Review of Pharmacology and Toxicology, 27(1): p. 399-427.
266. Chen Y.-T., et al., 2010. *Protective effects of fomepizole on 2-chloroethanol toxicity*. Human & Experimental Toxicology, 29(6): p. 507-512.
267. Jones A.R. and Wells G., 1981. *The comparative metabolism of 2-bromoethanol and ethylene oxide in the rat*. Xenobiotica, 11(11): p. 763-770.
268. Tewson T.J. and Welch M.J., 1980. *Preparation and preliminary biodistribution of "no carrier added" fluorine 1-8 fluoroethanol*. Journal of Nuclear Medicine, 21(6): p. 559-564.
269. Wierchowski J., et al., 1997. *Fluorimetric detection of aldehyde dehydrogenase activity in human blood, saliva, and organ biopsies and kinetic differentiation between class I and class III isozymes*. Analytical Biochemistry, 245(1): p. 69-78.



270. Ditlow C.C., et al., 1984. *Physical and enzymic properties of a class II alcohol dehydrogenase isozyme of human liver:  $\pi$ -ADH*. *Biochemistry*, 23(26): p. 6363-6368.
271. Pietruszko R., et al., 1973. *Comparison of substrate specificity of alcohol dehydrogenases from human liver, horse liver, and yeast towards saturated and 2-enoic alcohols and aldehydes*. *Archives of Biochemistry and Biophysics*, 159(1): p. 50-60.
272. Dickinson F.M. and Dalziel K., 1967. *The specificities and configurations of ternary complexes of yeast and liver alcohol dehydrogenases*. *Biochem. J.*, 104(1): p. 165-172.
273. Wagner F.W., et al., 1984. *Physical and enzymatic properties of a class III isozyme of human liver alcohol dehydrogenase:  $\chi$ -ADH*. *Biochemistry*, 23(10): p. 2193-2199.
274. Edwards J.E., et al., 2005. *The metabolism of nonane, a JP-8 jet fuel component, by human liver microsomes, P450 isoforms and alcohol dehydrogenase and inhibition of human P450 isoforms by JP-8*. *Chemico-Biological Interactions*, 151(3): p. 203-211.
275. Stone C.L., et al., 1989. *Stereospecific oxidation of secondary alcohols by human alcohol dehydrogenases*. *Journal of Biological Chemistry*, 264(19): p. 11112-11116.
276. Kemper R.A. and Elfarra A.A., 1996. *Oxidation of 3-butene-1,2-diol by alcohol dehydrogenase*. *Chemical Research in Toxicology*, 9(7): p. 1127-1134.
277. Rikans L.E., 1987. *The oxidation of acrolein by rat liver aldehyde dehydrogenases. Relation to allyl alcohol hepatotoxicity*. *Drug Metabolism and Disposition*, 15(3): p. 356-362.
278. Sprague C.L. and Elfarra A.A., 2003. *Detection of carboxylic acids and inhibition of hippuric acid formation in rats treated with 3-butene-1,2-diol, a major metabolite of 1,3-butadiene*. *Drug Metabolism and Disposition*, 31(8): p. 986-992.
279. Glatt H., et al., 2008. *Detoxification of promutagenic aldehydes derived from methylpyrenes by human aldehyde dehydrogenases ALDH2 and ALDH3A1*. *Archives of Biochemistry and Biophysics*, 477(2): p. 196-205.
280. Bosron W.F., et al., 1979. *Human liver  $\pi$ -alcohol dehydrogenase: Kinetic and molecular properties*. *Biochemistry*, 18(6): p. 1101-1105.
281. Bosron W.F., et al., 1983. *Kinetic and electrophoretic properties of native and recombined isoenzymes of human liver alcohol dehydrogenase*. *Biochemistry*, 22(8): p. 1852-1857.

282. Burnell J.C., et al., 1989. *Purification and steady-state kinetic characterization of human liver  $\beta_3\beta_3$  alcohol dehydrogenase*. Biochemistry, 28(17): p. 6810-6815.
283. Eklund H., et al., 1990. *Comparison of three classes of human liver alcohol dehydrogenase*. European Journal of Biochemistry, 193(2): p. 303-307.
284. Herrera E., et al., 1983. *Comparative kinetics of human and rat liver alcohol dehydrogenase*. Biochemical Society Transactions, 11: p. 729-730.
285. Juliá P., et al., 1987. *Characterization of three isoenzymes of rat alcohol dehydrogenase*. European Journal of Biochemistry, 162(1): p. 179-189.
286. Lange L.G., et al., 1976. *Human liver alcohol dehydrogenase: Purification, composition, and catalytic features*. Biochemistry, 15(21): p. 4687-4693.
287. Ryzewski C.N. and Pietruszko R., 1977. *Horse liver alcohol dehydrogenase SS: Purification and characterization of the homogeneous isoenzyme*. Archives of Biochemistry and Biophysics, 183(1): p. 73-82.
288. Wagner F.W., et al., 1983. *Kinetic properties of human liver alcohol dehydrogenase: Oxidation of alcohols by class I isoenzymes*. Biochemistry, 22(8): p. 1857-1863.
289. Wagner F.W., et al., 1984. *Physical and enzymatic properties of a class III isozyme of human liver alcohol dehydrogenase:  $\chi$ -ADH*. Biochemistry, 23(10): p. 2193-2199.
290. Ambroziak W. and Pietruszko R., 1991. *Human aldehyde dehydrogenase. Activity with aldehyde metabolites of monoamines, diamines, and polyamines*. Journal of Biological Chemistry, 266(20): p. 13011-13018.
291. Chern M.K. and Pietruszko R., 1995. *Human aldehyde dehydrogenase E3 isozyme is a betaine aldehyde dehydrogenase*. Biochemical and Biophysical Research Communications, 213(2): p. 561-568.
292. Eckfeldt J., et al., 1976. *Horse liver aldehyde dehydrogenase. Purification and characterization of two isozymes*. Journal of Biological Chemistry, 251(1): p. 236-240.
293. Farrés J., et al., 1994. *Effects of changing glutamate 487 to lysine in rat and human liver mitochondrial aldehyde dehydrogenase. A model to study human (Oriental type) class 2 aldehyde dehydrogenase*. Journal of Biological Chemistry, 269(19): p. 13854-13860.
294. Greenfield N.J. and Pietruszko R., 1977. *Two aldehyde dehydrogenases from human liver. Isolation via affinity chromatography and characterization of the isozymes*. Biochimica et Biophysica Acta, Enzymology, 483(1): p. 35-45.

295. Han I.O. and Joo C.N., 1991. *Purification and characterization of the rat liver mitochondrial aldehyde dehydrogenase*. Korean Biochemical Journal, 24(4): p. 353-360.
296. Izaguirre G., et al., 1998. *Methylglyoxal as substrate and inhibitor of human aldehyde dehydrogenase: Comparison of kinetic properties among the three isozymes*. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 119(4): p. 747-754.
297. Kurys G., et al., 1989. *Human aldehyde dehydrogenase. Purification and characterization of a third isozyme with low Km for gamma-aminobutyraldehyde*. Journal of Biological Chemistry, 264(8): p. 4715-4721.
298. Lindahl R. and Evces S., 1984. *Rat liver aldehyde dehydrogenase. I. Isolation and characterization of four high Km normal liver isozymes*. Journal of Biological Chemistry, 259(19): p. 11986-11990.
299. Martini R. and Murray M., 1996. *Characterization of the in vivo inhibition of rat hepatic microsomal aldehyde dehydrogenase activity by metyrapone*. Biochemical Pharmacology, 51(9): p. 1187-1193.
300. Rashkovetsky L.G., et al., 1994. *Human liver aldehyde dehydrogenases: New method of purification of the major mitochondrial and cytosolic enzymes and re-evaluation of their kinetic properties*. Biochimica et Biophysica Acta, Protein Structure and Molecular Enzymology, 1205(2): p. 301-307.
301. Rietveld E.C., et al., 1987. *Substituent effects during the rat liver aldehyde dehydrogenase catalyzed oxidation of aromatic aldehydes*. Biochimica et Biophysica Acta, Protein Structure and Molecular Enzymology, 914(2): p. 162-169.
302. Rout U.K. and Weiner H., 1994. *Involvement of serine 74 in the enzyme-coenzyme interaction of rat liver mitochondrial aldehyde dehydrogenase*. Biochemistry, 33(30): p. 8955-8961.
303. Siew C., et al., 1976. *Localization and characteristics of rat liver mitochondrial aldehyde dehydrogenases*. Archives of Biochemistry and Biophysics, 176(2): p. 638-649.
304. Cashman J.R. and Ziegler D.M., 1986. *Contribution of N-oxygenation to the metabolism of MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) by various liver preparations*. Molecular Pharmacology, 29(2): p. 163-167.
305. Light D.R., et al., 1982. *Studies on the chirality of sulfoxidation catalyzed by bacterial flavoenzyme cyclohexanone monooxygenase and hog liver FAD-containing monooxygenase*. Biochemistry, 21(10): p. 2490-2498.
306. McManus M.E., et al., 1983. *Guanethidine N-oxide formation as a measure of cellular flavin-containing monooxygenase activity*. Biochemical and Biophysical Research Communications, 112(2): p. 437-443.

307. Nnane I.P. and Damani L.A., 2003. *Sulphoxidation of ethyl methyl sulphide, 4-chlorophenyl methyl sulphide and diphenyl sulphide by purified pig liver flavin-containing monooxygenase*. *Xenobiotica*, 33(1): p. 83-91.
308. Poulsen L.L. and Ziegler D.M., 1977. *Microsomal mixed-function oxidase-dependent renaturation of reduced ribonuclease*. *Archives of Biochemistry and Biophysics*, 183(2): p. 563-570.
309. Poulsen L.L., et al., 1974. *S-oxidation of thioureylenes catalyzed by a microsomal flavoprotein mixed-function oxidase*. *Biochemical Pharmacology*, 23(24): p. 3431-3440.
310. Prough R.A., 1973. *The N-oxidation of alkylhydrazines catalyzed by the microsomal mixed-function amine oxidase*. *Archives of Biochemistry and Biophysics*, 158(1): p. 442-444.
311. Prough R.A., et al., 1981. *The oxidation of hydrazine derivatives catalyzed by the purified liver microsomal FAD-containing monooxygenase*. *Journal of Biological Chemistry*, 256(9): p. 4178-4184.
312. Sabourin P.J. and Hodgson E., 1984. *Characterization of the purified microsomal FAD-containing monooxygenase from mouse and pig liver*. *Chemico-Biological Interactions*, 51(2): p. 125-139.
313. Smyser B.P., et al., 1985. *Oxidation of pesticides by purified microsomal FAD-containing monooxygenase from mouse and pig liver*. *Pesticide Biochemistry and Physiology*, 24(3): p. 368-374.
314. Sofer S.S. and Ziegler D.M., 1978. *Microsomal mixed-function amine oxidase. Oxidation products of piperazine-substituted phenothiazine drugs*. *Drug Metabolism and Disposition*, 6(3): p. 232-239.
315. Tynes R.E. and Hodgson E., 1985a. *Magnitude of involvement of the mammalian flavin-containing monooxygenase in the microsomal oxidation of pesticides*. *Journal of Agricultural and Food Chemistry*, 33(3): p. 471-479.
316. Tynes R.E. and Hodgson E., 1985b. *Catalytic activity and substrate specificity of the flavin-containing monooxygenase in microsomal systems: Characterization of the hepatic, pulmonary and renal enzymes of the mouse, rabbit, and rat*. *Archives of Biochemistry and Biophysics*, 240(1): p. 77-93.
317. Venkatesh K., et al., 1991. *The flavin-containing monooxygenase of mouse kidney: A comparison with the liver enzyme*. *Biochemical Pharmacology*, 42(7): p. 1411-1420.
318. Wu R.F. and Ichikawa Y., 1994. *Characteristic properties and kinetic analysis with neurotoxins of porcine FAD-containing monooxygenase*. *Biochimica et Biophysica Acta, Protein Structure and Molecular Enzymology*, 1208(2): p. 204-210.

319. Wu R.F., et al., 1992. *Neurotoxins: 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine, 1,2,3,4-tetrahydroisoquinoline and 1-methyl-6,7-dihydroxy-tetrahydroisoquinoline as substrates for FAD-containing monooxygenase of porcine liver microsomes*. *Biochemical Pharmacology*, 44(10): p. 2079-2081.
320. Wu R.F., et al., 2004. *Porcine FAD-containing monooxygenase metabolizes lidocaine, bupivacaine and propranolol in vitro*. *Life Sciences*, 75(8): p. 1011-1019.
321. Bessems J.G.M., et al., 1996. *Rat liver microsomal cytochrome P450-dependent oxidation of 3,5-disubstituted analogues of paracetamol*. *Xenobiotica*, 26(6): p. 647-666.
322. Blake R.C. and Coon M.J., 1981. *On the mechanism of action of cytochrome P-450. Evaluation of homolytic and heterolytic mechanisms of oxygen-oxygen bond cleavage during substrate hydroxylation by peroxides*. *Journal of Biological Chemistry*, 256(23): p. 12127-12133.
323. Galliani G., et al., 1986. *The rate of N-demethylation of N,N-dimethylanilines and N-methylanilines by rat-liver microsomes is related to their first ionization potential, their lipophilicity and to a steric bulk factor*. *Xenobiotica*, 16(6): p. 511-517.
324. Macdonald T.L., et al., 1989. *Oxidation of substituted N,N-dimethylanilines by cytochrome P-450: Estimation of the effective oxidation-reduction potential of cytochrome P-450*. *Biochemistry*, 28(5): p. 2071-2077.
325. Martin Y.C. and Hansch C., 1971. *Influence of hydrophobic character on the relative rate of oxidation of drugs by rat liver microsomes*. *Journal of Medicinal Chemistry*, 14(9): p. 777-779.
326. Morgan E.T., et al., 1982. *Catalytic activity of cytochrome P-450 isozyme 3a isolated from liver microsomes of ethanol-treated rabbits. Oxidation of alcohols*. *Journal of Biological Chemistry*, 257(23): p. 13951-13957.
327. Peng H.M., et al., 1995. *Oxidative cleavage of esters and amides to carbonyl products by cytochrome P450*. *Archives of Biochemistry and Biophysics*, 318(2): p. 333-339.
328. Vaz A.D.N. and Coon M.J., 1994. *On the mechanism of action of cytochrome P450: Evaluation of hydrogen abstraction in oxygen-dependent alcohol oxidation*. *Biochemistry*, 33(21): p. 6442-6449.
329. Watanabe Y., et al., 1980. *Kinetic study on enzymatic s-oxygenation promoted by a reconstituted system with purified cytochrome P-450*. *Tetrahedron Letters*, 21(38): p. 3685-3688.
330. Watanabe Y., et al., 1982. *One electron transfer mechanism in the enzymatic oxygenation of sulfoxide to sulfone promoted by a reconstituted system with purified cytochrome p-450*. *Tetrahedron Letters*, 23(5): p. 533-536.

331. White R.E. and McCarthy M.-B., 1986. *Active site mechanics of liver microsomal cytochrome P-450*. Archives of Biochemistry and Biophysics, 246(1): p. 19-32.
332. Pelkonen O., et al., 2009. *Comparison of metabolic stability and metabolite identification of 55 ECVAM/ICCVAM validation compounds between human and rat liver homogenates and microsomes - a preliminary analysis*. ALTEX, 26: p. 214-222.
333. Rotroff D.M., et al., 2010. *Incorporating human dosimetry and exposure into high-throughput in vitro toxicity screening*. Toxicological Sciences, 117(2): p. 348-358.
334. Yamazaki H., et al., 2010. *Human blood concentrations of dichlorodiphenyltrichloroethane (DDT) extrapolated from metabolism in rats and humans and physiologically based pharmacokinetic modeling*. Journal of Health Science, 56(5): p. 566-575.
335. Parham F.M. and Portier C.J., 1998. *Using structural information to create physiologically based pharmacokinetic models for all polychlorinated biphenyls*. Toxicology and Applied Pharmacology, 151(1): p. 110-116.
336. Han X., et al., 2011. *Comparative metabolism of 1,2,3,3,3-pentafluoropropene in male and female mouse, rat, dog, and human liver microsomes and cytosol and male rat hepatocytes via oxidative dehalogenation and glutathione S-conjugation pathways*. Drug Metabolism and Disposition, 39(7): p. 1288-1293.
337. Boogaard P.J. and Bond J.A., 1996. *The role of hydrolysis in the detoxification of 1,2:3,4-diepoxybutane by human, rat, and mouse liver and lung in vitro*. Toxicology and Applied Pharmacology, 141(2): p. 617-627.
338. Wormhoudt L.W., et al., 1996. *Inter-individual variability in the oxidation of 1,2-dibromoethane: Use of heterologously expressed human cytochrome P450 and human liver microsomes*. Chemico-Biological Interactions, 101(3): p. 175-192.
339. Csanády G.A., et al., 1992. *Comparison of the biotransformation of 1,3-butadiene and its metabolite, butadiene monoepoxide, by hepatic and pulmonary tissues from humans, rats and mice*. Carcinogenesis, 13(7): p. 1143-1153.
340. Duescher R.J. and Elfarra A.A., 1994. *Human liver microsomes are efficient catalysts of 1,3-butadiene oxidation: Evidence for major roles by cytochromes P450 2A6 and 2E1*. Archives of Biochemistry and Biophysics, 311(2): p. 342-349.

341. Schnellmann R.G., et al., 1983. *Metabolism of 2,2',3,3',6,6'-hexachlorobiphenyl and 2,2',4,4',5,5'-hexachlorobiphenyl by human hepatic microsomes*. *Biochemical Pharmacology*, 32(21): p. 3233-3239.
342. Erratico C.A., et al., 2012. *Oxidative metabolism of BDE-99 by human liver microsomes: Predominant role of CYP2B6*. *Toxicological Sciences*, 129(2): p. 280-292.
343. Erratico C.A., et al., 2013. *Biotransformation of 2,2',4,4'-tetrabromodiphenyl ether (BDE-47) by human liver microsomes: Identification of cytochrome P450 2B6 as the major enzyme involved*. *Chemical Research in Toxicology*, 26(5): p. 721-731.
344. Feo M.L., et al., 2013. *Biotransformation of BDE-47 to potentially toxic metabolites is predominantly mediated by human CYP2B6*. *Environmental Health Perspectives*, 121(4): p. 440-446.
345. Roberts S.C., et al., 2012. *In vitro metabolism of the brominated flame retardants 2-ethylhexyl-2,3,4,5-tetrabromobenzoate (TBB) and bis(2-ethylhexyl) 2,3,4,5-tetrabromophthalate (TBPH) in human and rat tissues*. *Chemical Research in Toxicology*, 25(7): p. 1435-1441.
346. Schnellmann R.G., et al., 1984. *The hydroxylation, dechlorination, and glucuronidation of 4,4'-dichlorobiphenyl (4-DCB) by human hepatic microsomes*. *Biochemical Pharmacology*, 33(21): p. 3503-3509.
347. Nabb D.L., et al., 2007. *In vitro metabolism of 8-2 fluorotelomer alcohol: Interspecies comparisons and metabolic pathway refinement*. *Toxicological Sciences*, 100(2): p. 333-344.
348. Kasai N., et al., 2004. *Sequential metabolism of 2,3,7-trichlorodibenzo-p-dioxin (2,3,7-tricdd) by cytochrome P450 and UDP-glucuronosyltransferase in human liver microsomes*. *Drug Metabolism and Disposition*, 32(8): p. 870-875.
349. Obach R.S., 1999. *Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: An examination of in vitro half-life approach and nonspecific binding to microsomes*. *Drug Metabolism and Disposition*, 27(11): p. 1350-1359.
350. Abass K., et al., 2014a. *Human variation and CYP enzyme contribution in benfuracarb metabolism in human in vitro hepatic models*. *Toxicology Letters*, 224(2): p. 300-309.
351. Abass K., et al., 2014b. *Comparative metabolism of benfuracarb in in vitro mammalian hepatic microsomes model and its implications for chemical risk assessment*. *Toxicology Letters*, 224(2): p. 290-299.
352. Nedelcheva V., et al., 1999. *Metabolism of benzene in human liver microsomes: Individual variations in relation to CYP2E1 expression*. *Archives of Toxicology*, 73(1): p. 33-40.

353. Crowell S.R., et al., 2014. *In vitro* metabolism of benzo[a]pyrene and dibenzo[def,p]chrysene in rodent and human hepatic microsomes. *Toxicology Letters*, 228(1): p. 48-55.
354. Himmelstein M.W., et al., 2004. *Kinetic modeling of  $\beta$ -chloroprene metabolism: I. In vitro rates in liver and lung tissue fractions from mice, rats, hamsters, and humans*. *Toxicological Sciences*, 79(1): p. 18-27.
355. Hanioka N., et al., 2008. *Human UDP-glucuronosyltransferase isoforms involved in bisphenol A glucuronidation*. *Chemosphere*, 74(1): p. 33-36.
356. Mazur C.S., et al., 2010. *Differences between human and rat intestinal and hepatic bisphenol a glucuronidation and the influence of alamethicin on in vitro kinetic measurements*. *Drug Metabolism and Disposition*, 38(12): p. 2232-2238.
357. Cabaton N., et al., 2008. *Biotransformation of bisphenol F by human and rat liver subcellular fractions*. *Toxicology in Vitro*, 22(7): p. 1697-1704.
358. Kerger B.D., et al., 1988. *Comparison of human and mouse liver microsomal metabolism of bromobenzene and chlorobenzene to 2- and 4-halophenols*. *Drug Metabolism and Disposition*, 16(5): p. 672-677.
359. Krause R.J. and Elfarra A.A., 1997. *Oxidation of butadiene monoxide to meso- and ( $\pm$ )-diepoxybutane by cDNA-expressed human cytochrome P450s and by mouse, rat, and human liver microsomes: Evidence for preferential hydration of meso-diepoxybutane in rat and human liver microsomes*. *Archives of Biochemistry and Biophysics*, 337(2): p. 176-184.
360. Kreuzer P.E., et al., 1991. *Enzyme specific kinetics of 1,2-epoxybutene-3 in microsomes and cytosol from livers of mouse, rat, and man*. *Archives of Toxicology*, 65(1): p. 59-67.
361. Tang J., et al., 2002. *In vitro* metabolism of carbaryl by human cytochrome P450 and its inhibition by chlorpyrifos. *Chemico-Biological Interactions*, 141(3): p. 229-241.
362. Usmani K.A., et al., 2004a. *In vitro* metabolism of carbofuran by human, mouse, and rat cytochrome P450 and interactions with chlorpyrifos, testosterone, and estradiol. *Chemico-Biological Interactions*, 150(3): p. 221-232.
363. Abass K., et al., 2009. *Metabolism of carbosulfan. I. Species differences in the in vitro biotransformation by mammalian hepatic microsomes including human*. *Chemico-Biological Interactions*, 181(2): p. 210-219.
364. Foxenberg R.J., et al., 2007. *Human hepatic cytochrome P450-specific metabolism of parathion and chlorpyrifos*. *Drug Metabolism and Disposition*, 35(2): p. 189-193.



365. Sams C., et al., 2004. *Biotransformation of chlorpyrifos and diazinon by human liver microsomes and recombinant human cytochrome P450s (CYP)*. *Xenobiotica*, 34(10): p. 861-873.
366. Smith J.N., et al., 2011. *In vitro age-dependent enzymatic metabolism of chlorpyrifos and chlorpyrifos-oxon in human hepatic microsomes and chlorpyrifos-oxon in plasma*. *Drug Metabolism and Disposition*, 39(8): p. 1353-1362.
367. Tang J., et al., 2001. *Metabolism of chlorpyrifos by human cytochrome P450 isoforms and human, mouse, and rat liver microsomes*. *Drug Metabolism and Disposition*, 29(9): p. 1201-1204.
368. Godin S.J., et al., 2006. *Species differences in the in vitro metabolism of deltamethrin and esfenvalerate: Differential oxidative and hydrolytic metabolism by humans and rats*. *Drug Metabolism and Disposition*, 34(10): p. 1764-1771.
369. Ellison C.A., et al., 2012. *Human hepatic cytochrome P450-specific metabolism of the organophosphorus pesticides methyl parathion and diazinon*. *Drug Metabolism and Disposition*, 40(1): p. 1-5.
370. Mráz J., et al., 1993. *Investigation of the mechanistic basis of N,N-dimethylformamide toxicity. Metabolism of N,N-dimethylformamide and its deuterated isotopomers by cytochrome P450 2E1*. *Chemical Research in Toxicology*, 6(2): p. 197-207.
371. Usmani K.A., et al., 2004b. *In vitro sulfoxidation of thioether compounds by human cytochrome P450 and flavin-containing monooxygenase isoforms with particular reference to the CYP2C subfamily*. *Drug Metabolism and Disposition*, 32(3): p. 333-339.
372. Abass K., et al., 2007b. *Characterization of diuron N-demethylation by mammalian hepatic microsomes and cDNA-expressed human cytochrome P450 enzymes*. *Drug Metabolism and Disposition*, 35(9): p. 1634-1641.
373. Casabar R.C.T., et al., 2006. *Metabolism of endosulfan- $\alpha$  by human liver microsomes and its utility as a simultaneous in vitro probe for CYP2B6 and CYP3A4*. *Drug Metabolism and Disposition*, 34(10): p. 1779-1785.
374. Tang J., et al., 2004. *In vitro metabolism of fipronil by human and rat cytochrome P450 and its interactions with testosterone and diazepam*. *Chemico-Biological Interactions*, 147(3): p. 319-329.
375. Green T., et al., 2003. *Assessing the health risks following environmental exposure to hexachlorobutadiene*. *Toxicology Letters*, 138(1-2): p. 63-73.
376. Hong J.-Y., et al., 1999. *Metabolism of methyl tert-butyl ether and other gasoline ethers by human liver microsomes and heterologously expressed*

*human cytochromes P450: Identification of CYP2A6 as a major catalyst.* Toxicology and Applied Pharmacology, 160(1): p. 43-48.

377. Reitz R.H., et al., 1988. *Incorporation of in vitro enzyme data into the physiologically-based pharmacokinetic (PB-PK) model for methylene chloride: implications for risk assessment.* Toxicology Letters, 43(1-3): p. 97-116.

378. Völkel W., et al., 1999. *Slow oxidation of acetoxime and methylethyl ketoxime to the corresponding nitronates and hydroxy nitronates by liver microsomes from rats, mice, and humans.* Toxicological Sciences, 47(2): p. 144-150.

379. Jewell W.T. and Miller M.G., 1999. *Comparison of human and rat metabolism of molinate in liver microsomes and slices.* Drug Metabolism and Disposition, 27(7): p. 842-847.

380. Cho T.M., et al., 2006. *In vitro metabolism of naphthalene by human liver microsomal cytochrome P450 enzymes.* Drug Metabolism and Disposition, 34(1): p. 176-183.

381. Xu L., et al., 2006. *N-glucuronidation of perfluorooctanesulfonamide by human, rat, dog, and monkey liver microsomes and by expressed rat and human UDP-glucuronosyltransferases.* Drug Metabolism and Disposition, 34(8): p. 1406-1410.

382. Abass K., et al., 2007a. *In vitro metabolism and interaction of profenofos by human, mouse and rat liver preparations.* Pesticide Biochemistry and Physiology, 87(3): p. 238-247.

383. Dadson O.A., et al., 2013. *Metabolism of profenofos to 4-bromo-2-chlorophenol, a specific and sensitive exposure biomarker.* Toxicology, 306: p. 35-39.

384. Faller T.H., et al., 2001. *Kinetics of propylene oxide metabolism in microsomes and cytosol of different organs from mouse, rat, and humans.* Toxicology and Applied Pharmacology, 172(1): p. 62-74.

385. Mendrala A.L., et al., 1993. *In vitro kinetics of styrene and styrene oxide metabolism in rat, mouse, and human.* Archives of Toxicology, 67(1): p. 18-27.

386. Lipscomb J.C., et al., 1997. *Cytochrome P450-dependent metabolism of trichloroethylene: Interindividual differences in humans.* Toxicology and Applied Pharmacology, 142(2): p. 311-318.

# Summary

## Quantifying biotransformation of xenobiotics in mammals

Biotransformation is one of the processes that influence the bioaccumulation of chemicals by decreasing the concentration of chemicals in an organism. In order to be metabolised, a chemical needs to bind to an enzyme and then a catalytic reaction takes place. Compounds are usually transformed into more hydrophilic metabolites, which are more easily eliminated from the organism. Predicting the biotransformation rate of a chemical is, however, a difficult task due to the specific action of metabolism, which depends on the chemical and the enzymes involved.

The aim of this thesis was to develop models for the prediction of biotransformation of xenobiotics (pharmaceuticals and environmental pollutants) in mammals based on their chemical properties. The relationships between metabolic activity and chemical structure were performed for *in vitro* systems representing different levels of biological organization (i.e. isolated enzymes, hepatocytes and microsomes). The mechanisms underlying metabolism were investigated starting from the enzyme level. The focus was on the liver metabolism in mammals mediated by four important oxidising enzymes: alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), flavin-containing monooxygenase (FMO) and cytochrome P450 (CYP) enzymes. Different types of descriptors were used in the model development, including the octanol-water partitioning coefficient ( $K_{ow}$ ), mechanistic descriptors and theoretical descriptors.

In Chapter 2, the change in hydrophobicity, expressed as  $K_{ow}$ , was quantified for organic pollutants undergoing various biotransformation reactions in mammals. The  $K_{ow}$  values of a selected dataset of parent compounds were compared with the  $K_{ow}$  of their first metabolites following oxidation reactions catalysed by CYP, ADH and ALDH. The  $K_{ow}$  decreased up to two orders of magnitude, depending on the metabolic pathway. For reactions mediated by CYP, the decrease in  $K_{ow}$  was one order of magnitude for hydroxylated and epoxidated compounds and two orders of magnitude for dihydroxylated and sulfoxidated xenobiotics. In contrast, no significant change in hydrophobicity was observed for compounds N-hydroxylated by CYP and for alcohols and aldehydes metabolised by ADH and ALDH. These relationships estimate the extent to which the elimination of pollutants is increased by biotransformation. Thus, the quantification of the  $K_{ow}$  reduction might be considered as a first step in predicting biotransformation rates, but further studies are needed to investigate the feasibility of this approach.

In Chapter 3, binding affinity, expressed as  $1/K_m$  was related to compound hydrophobicity, expressed as  $K_{ow}$ , for compounds oxidised by ADH, ALDH, FMO and CYP enzymes. For all regressions,  $1/K_m$  increased with compound  $K_{ow}$ , which can be understood from the tendency to biotransform hydrophobic compounds into more polar, thus more easily excretable metabolites. Hydrophobicity was relevant to the binding of most of the substrate classes of ADH, ALDH and CYP. The resulting slopes had 95% Confidence Intervals covering the value of 0.6, typically noted in protein-water distribution regressions on the basis of  $K_{ow}$ . If weak interactions are dominant, the partitioning of organic chemicals over various phases is governed by hydrophobicity and polarity, thus it can be related to compound  $K_{ow}$ . A reduced slope (0.2–0.3) was found for FMO: this may be due to a different reaction mechanism involving a nucleophilic attack, which is a strong interaction thus it cannot be explained with compound  $K_{ow}$ .

In Chapter 4, models were developed to better understand how binding affinity ( $1/K_m$ ) and maximum reaction rate ( $V_{max}$ ) for substrates of ADH, ALDH, FMO and CYP in mammals relate to partitioning, geometric characteristics and electronic properties of the substrates. The explained variance of the models varied between 20% and 70% and was larger for  $1/K_m$  than for  $V_{max}$ . The increase of  $1/K_m$  with compound hydrophobicity and size suggests that weak interactions are important, e.g. by substrate binding via desolvation processes. The importance of electronic factors for  $1/K_m$  was described in relation to the catalytic mechanism of the enzymes.  $V_{max}$  was particularly influenced by electronic properties, such as dipole moment and energy of the lowest unoccupied molecular orbital. This can be explained by the nature of the catalysis, characterised by the cleavage and formation of covalent or ionic bonds (strong interactions).

In Chapter 5, predictive models were developed for the enzymatic constants using theoretical descriptors. 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 models were between 50% and 80% and the predictive abilities ( $R^2_{ext}$ ) between 50% and 60%, except for the  $V_{max}$  model 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  $1/K_m$  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.

Successively,  $1/K_m$  and  $V_{max}$  values were also collected for whole liver cells and sub-cellular fractions (hepatocytes and microsomes) to build models predicting *in vitro* clearance ( $CL_{INT}$ , i.e.  $V_{max}/K_m$ ) for humans. In Chapter 6, multiple linear models were built and validated selecting at most 6 predictors from a pool of over 2000 potential molecular descriptors. For the hepatocytes model, the explained variance ( $R^2_{adj}$ ) was 67% and the predictive ability ( $R^2_{ext}$ ) was 62%. For the microsomes model,  $R^2_{adj}$  was 50% and  $R^2_{ext}$  30%. For both liver assays, the most important descriptor relates to electronic properties of the compound. Functional groups of fragments were useful to identify specific compounds that have a deviating reaction rate compared to the others, such as Polychlorobiphenyls (PCBs) and organic amides which were poorly metabolised.

Finally, in Chapter 7 the advantages and disadvantages of the different types of descriptors and levels of biological organization were discussed. While the models for individual enzymes were helpful to interpret metabolic processes, their application to risk assessment is limited. Instead, the most promising results were obtained with human hepatocytes. In addition, a general scheme to perform *in vitro-in vivo* extrapolations (*ivive*) was proposed and evaluated. The performances of the models were, however, limited by the reliability of the *in vitro* assay systems. The models can potentially be improved when more *in vitro* data become available from standardised experiments. In addition, the *ivive* method needs to be validated on a wide array of chemicals, yet it could be useful for a first estimate of  $k_m$  in a weight of evidence approach.

# Samenvatting

## Kwantificeren van biotransformatie van lichaamsvreemde stoffen in zoogdieren

Biotransformatie is één van de processen die de bioaccumulatie van chemische stoffen beïnvloeden door de concentratie in organisme te verminderen. Om gemetaboliseerd te worden moet een stof binden aan een enzym waarna een katalytische reactie plaatsvindt. Stoffen worden meestal omgezet naar meer wateroplosbare metabolieten, die makkelijker geëlimineerd worden door het organisme. Het voorspellen van de biotransformatie snelheid is echter lastig vanwege de specifieke werking die afhangt van de stof en de betrokken enzymen.

De doelstelling van dit proefschrift was het ontwikkelen van modellen om de biotransformatie van lichaamsvreemde stoffen (xenobiotica, t.w. medicijnen en milieu-verontreinigingen) in zoogdieren te voorspellen op basis van hun chemische eigenschappen. De relaties tussen de metabole activiteit en de chemische structuur werden gelegd voor *in vitro* systemen die verschillende niveau's van biologische organisatie (d.w.z. geïsoleerde enzymen, levercellen en microsomen) representeren. De onderliggende mechanismen werden onderzocht, allereerst op het niveau van enzymen. The focus lag op afbraak in de lever door vier belangrijke oxiderende enzymen: alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), flavin-containing monooxygenase (FMO) en cytochrome P450 (CYP). Verschillende typen descriptoren werden gebruikt in de modelontwikkeling, waaronder de octanol-water partitie coefficient ( $K_{ow}$ ), alsook mechanistische en theoretische descriptoren.

In Hoofdstuk 2 is de verandering in de hydrofobiciteit, uitgedrukt in  $K_{ow}$ , gekwantificeerd voor organische stoffen die verschillende biotransformatie reacties ondergaan in zoogdieren. The  $K_{ow}$  waarden van een dataset van moederstoffen is vergeleken met de  $K_{ow}$  van hun eerste metabolieten volgend op oxidatie reacties, gecatalyseerd door CYP, ADH en ALDH. De  $K_{ow}$  nam tot twee ordes van grootte af, afhankelijk van de metabole route. Voor CYP gemedieerde reacties, was de afname in  $K_{ow}$  één orde van grootte voor gehydroxyleerde en geëpoxydeerde stoffen en twee ordes van grootte voor gedihydroxyleerde and gesulfoneerde xenobiotica. Daarentegen was de afname in hydrofobiciteit niet significant voor door CYP N-gehydroxyleerde stoffen en door ADH en ALDH gemetaboliseerde alcoholen en aldehydes. Met deze relaties kan de mate waarin eliminatie verhoogd is door biotransformatie geschat worden, maar vervolgstudies zijn nodig om de haalbaarheid van deze benadering te onderzoeken.

In Hoofdstuk 3 is de bindingsaffiniteit, uitgedrukt als  $1/K_m$  gerelateerd aan de hydrophobicity, uitgedrukt als  $K_{ow}$ , voor stoffen die worden geoxideerd door ADH, ALDH, FMO en CYP enzymen. In alle regressies nam  $1/K_m$  toe met de  $K_{ow}$  van de stof, hetgeen kan worden verklaard door de neiging om hydrofobe verbindingen om te zetten in meer polaire, en dus makkelijker uit te scheiden metabolieten. Hydrophobicity was relevant voor de binding van de meeste substraat klassen voor ADH, ALDH en CYP. De resulterende hellingen hadden 95% betrouwbaarheidsintervallen met daarin 0.6, de waarde die vaak wordt waargenomen in eiwit-water verdeling regressies op basis van de  $K_{ow}$ . Als zwakke interacties dominant zijn, wordt de verdeling van organische chemicaliën over verschillende fase bepaald door hydrophobiciteit en polariteit, zodat het gerelateerd kan worden aan de  $K_{ow}$  van de stof. De helling voor FMO was lager (0.2-0.3) waarschijnlijk omdat het reactie mechanisme anders is, met een nucleofiele aanval en dus een sterke interactie die niet met  $K_{ow}$  beschreven kan worden.

In Hoofdstuk 4 zijn modellen ontwikkeld om beter te begrijpen hoe bindingsaffiniteit ( $1/K_m$ ) en maximum reactie snelheid ( $V_{max}$ ) voor substraten van ADH, ALDH, FMO en CYP in zoogdieren gerelateerd zijn aan partitie, geometrische en electronische eigenschappen van substraten. De verklaarde variantie van de modellen varieerde tussen de 20% en 70% en was groter voor  $1/K_m$  dan voor  $V_{max}$ . De toename van  $1/K_m$  met de hydrophobiciteit en de grootte van de stof suggereert dat zwakke interacties bijvoorbeeld substraat binding via desolvatie, belangrijk zijn.  $V_{max}$  werd vooral beïnvloed door electronische eigenschappen, zoals dipoolmoment en de energie van de laagste onbezette moleculaire schil. Dit kan worden verklaard door de aard van de katalyse, gekarakteriseerd door de splitsing en vorming van covalente en ionbindingen (sterke interacties).

In Hoofdstuk 5, zijn voorspellende modellen ontwikkeld voor enzymatische constanten op basis van theoretische descriptoren. Een genetisch algoritme is toegepast om maximaal zes predictoren te selecteren uit een set van meer dan 2000 potentiële descriptoren, waarbij steeds twee-derde van de xenobiotica in elke enzym klasse werden gebruikt. De multiple lineaire modellen zijn daarna getoetst in een cross-validation met het resterende deel van de stoffen. De verklaarde variantie ( $R^2_{adj}$ ) van de modellen was tussen 50% en 80% en het voorspellend vermogen ( $R^2_{ext}$ ) tussen 50% en 60%, met uitzondering van die voor de  $V_{max}$  van MFO ( $R^2_{adj} < 30\%$ ,  $R^2_{ext} < 30\%$ ). The  $V_{max}$  waarden van FMO waren onafhankelijk van de chemische structuur van het substraat omdat de snelheids-beperkende stap van de katalytische cyclus voor de oxidatie ligt. Voor de andere enzymen werd  $V_{max}$  vooral bepaald door functionele groepen of fragmenten en door electronische eigenschappen vanwege de sterke en specifieke interacties bij de betrokken reacties. De meest relevante

predictoren voor  $1/K_m$  waren functionele groepen en fragmenten voor enzymen die specifieke stoffen metaboliseren (ADH, ALDH en FMO) en grootte en vorm eigenschappen voor CYP, waarschijnlijk vanwege de brede substraat specificiteit van CYP enzymen.

Vervolgens werden ook  $1/K_m$  and  $V_{max}$  waarden verzameld voor complete levercellen en sub-cellulaire fracties van levercellen en microsomen om humane modellen voor *in vitro* clearance ( $CL_{INT}$ , i.e.  $V_{max}/K_m$ ) te bouwen. In Hoofdstuk 6 zijn multiële lineaire modellen gebouwd en gevalideerd waarbij opnieuw maximaal 6 predictoren werden geselecteerd uit een set van 2000 potentiële descriptoren. Voor het levercel model was de verklaarde variantie ( $R^2_{adj}$ ) 67% and het voorspellend vermogen ( $R^2_{ext}$ ) 62%. Voor het microsoom model was  $R^2_{adj}$  50% en  $R^2_{ext}$  30%. De belangrijkste descriptoren voor beide lever testen waren gerelateerd aan de electronische eigenschappen van de stof. Functionele groepen van fragmenten bleken bruikbaar om specifieke stoffen met een afwijkende reactiesnelheid te identificeren, bijvoorbeeld bij slecht afbreekbare polychloorbiphenylen (PCBs) en organische amides.

Tenslotte zijn in Hoofdstuk 7 de voor- en nadelen van verschillende typen descriptoren op verschillende niveau's van biologische organisatie bediscussieerd. Hoewel de modellen voor individuele enzymens bruikbaar waren om metabole processen te interpreteren is hun toepassing in de risicobeoordeling beperkt. Daarentegen zijn veelbelovende resultaten verkregen voor humane levercellen. Bovendien is een algemeen schema afgeleid en geëvalueerd voor *in vitro* – *in vivo* extrapolatie. De prestaties van de modellen zijn echter beperkt door de betrouwbaarheid van de *in vitro* assay systemen. De modellen kunnen verbeterd worden wanneer meer *in vitro* data uit gestandaardiseerde experimenten beschikbaar komen. Daarnaast moet de *in vitro* - *in vivo* extrapolatie getest worden op een breed spectrum aan stoffen. Deze benadering kan geschikt zijn voor een eerste schatting van  $k_m$  in een “weight of evidence approach”.



# About the author

## Curriculum vitae

Alessandra Pirovano was born on 16 April 1984 in Melzo, Milano (Italy). She studied Environmental Sciences and Technologies at Bicocca University in Milano from 2004 until 2010. She focused on environmental chemistry, especially regarding fate and behaviour of pollutants. She obtained her BSc degree in 2007 with a thesis on data analysis of emissions of chlorinated organic micropollutants from a secondary casting aluminium plant. During her MSc, she attended a summer school in chemical/biochemical unit operation laboratory at the Technical University of Denmark (DTU) in Lyngby (Denmark). For her MSc thesis, she investigated the kinetics and mechanisms of formation and destruction of Polychlorinated Dibenzo-*p*-Dioxins and DibenzoFurans (PCDD/Fs). After her graduation in 2010, Alessandra started working as a junior researcher at the Department of Environmental Science of Radboud University in Nijmegen, where she carried out the PhD research that resulted in this thesis. Her work was financed for three years by the ITN Marie Curie project ECO (Environmental ChemOinformatics). During this period, she made a one month research visit to the Helmholtz Zentrum München (Germany). Currently, she is working at the Environmental Chemicals Agency (ECHA) in Helsinki (Finland) as a scientific and administrative assistant.

## (To be) peer-reviewed journal articles

Pirovano A, Borile N, Hendriks AJ. 2012. A comparison of octanol-water partitioning between organic chemicals and their metabolites in mammals. *Chemosphere*, 88(8), 1036–1041.

Pirovano A, Huijbregts MAJ, Ragas AMJ, Hendriks AJ. 2012. Compound lipophilicity as a descriptor to predict binding affinity ( $1/K_m$ ) in mammals. *Environmental Science & Technology*, 46(9), 5168-5174.

Pirovano A, Huijbregts MAJ, Ragas AMJ, Veltman K, Hendriks AJ. 2014. Mechanistically-based QSARs to describe metabolic constants in mammals. *Alternatives to Laboratory Animals*, 42(1), 59–69.

Pirovano A, Brandmaier S, Huijbregts MAJ, Ragas AMJ, Veltman K, Hendriks AJ. 2015. The utilisation of structural descriptors to predict metabolic constants of xenobiotics in mammals. *Environmental Toxicology and Pharmacology*, 39(1), 247-258.

Pirovano A, Brandmaier S, Huijbregts MAJ, Ragas AMJ, Veltman K, Hendriks AJ. QSARs for estimating intrinsic hepatic clearance of organic chemicals in humans. Submitted.

# Acknowledgments

I would like to acknowledge the support and guidance of my supervisors Jan, Mark, Ad and Karin. Jan, although involved with many projects and head of department duties, you have always found time to discuss issues and quickly provide comments and feedback. Thank you for your patience and consistent encouragement. Mark, your systematic approach to science has been really helpful to this research work. Thank you also for those times when I was stuck or lost and pointlessly trying to figure out things by myself, you understood it and came to help. Ad, thank you for your insightful inputs regarding both biology and statistics issues, as well as for carefully reviewing the contents and form of the manuscripts. Karin, I have highly appreciated your ability to go the core of the topics and understand them, thank you for providing constructive comments and improvements.

I would also like to acknowledge the ITN ECO project for financing my research and trainings. Thanks to Eva Schlosser for taking care of us fellows and to Igor Tetko for hosting me at Helmholtz Zentrum. Thanks to my ECO fellows Stefan and Pantelis for being such a great company when I was in Munich. Stefan, thanks also for your contribution to the model development for two of the papers in this thesis.

In addition, I would like to thank all my colleagues at the Department of Environmental Science. Isabel, we started together and supported each other all the way through, which was a great relief. Huge thanks for sharing with me thoughts, journeys and free time. Thanks to Laura and Anastasia for good times and chats on work and life. Jon, Daan and Rik, you made the Department livelier. Thanks to all the above colleagues also for the fun trips we had together. Zoran, thanks for addressing all my questions on statistics or R issues and for finding the solution. Gina, thanks for your kind assistance. Thanks to Ligia, Lisette, Katja and all other people working at the Department for your help and company.

This experience abroad was valuable also because it gave me the possibility to meet lovely people that made my time in the Netherlands unforgettable. Pragnya, thanks for your important support, hospitality and for countless laughs and chats. Ba, grazie della tua generosità e entusiasmo, è sempre un gran piacere rivederti. Ruben, thanks for the interesting talks on life, music and movies. Thanks for all the good times in Nijmegen (and beyond) also to Albert, Tim, Tobias, Dorien, Paola, Marco, Melina, Clara and Leti.

Un ringraziamento anche agli amici di vecchia data, quelli che mi fanno sempre sentire a casa quando torno, non importa quanto tempo io stia via. In particolare grazie a quelli che sono venuti a trovarmi (Lara per esserci sempre, Fabio, Lau, Simo e Ste) e a quelli che preferiscono fare i viaggi nei posti caldi (Ste, Luca e Dome).

Il mio ringraziamento più grande va alla mia famiglia. Ai nonni, per avermi insegnato cose che non si trovano sui libri. A mamma e papà, per avermi sempre supportato in ogni mia scelta, indipendentemente da quanto lontano fosse da casa. A Miriam e Simone, perché so che posso sempre contare su di voi, indipendentemente da quanto lontano io sia da casa. Ad Alberto, per il costante supporto e vicinanza nonostante la distanza e per incoraggiarmi ad andare avanti.