# Compound Lipophilicity as a Descriptor to Predict Binding Affinity (1/K-m) in Mammals

**ARTICLE** *in* ENVIRONMENTAL SCIENCE & TECHNOLOGY · APRIL 2012

Impact Factor: 5.33 · DOI: 10.1021/es204506g · Source: PubMed



**4 AUTHORS**, INCLUDING:

Alessandra Pirovano
Radboud University Nijmegen
**5** PUBLICATIONS **8** CITATIONS
SEE PROFILE

Mark A J Huijbregts
Radboud University Nijmegen
**239** PUBLICATIONS **5,797** CITATIONS
SEE PROFILE

Jan Hendriks
Radboud University Nijmegen
**208** PUBLICATIONS **3,587** CITATIONS
SEE PROFILE

Environmental Science & Technology

Article

pubs.acs.org/est

# Compound Lipophilicity as a Descriptor to Predict Binding Affinity ($1/K_m$) in Mammals

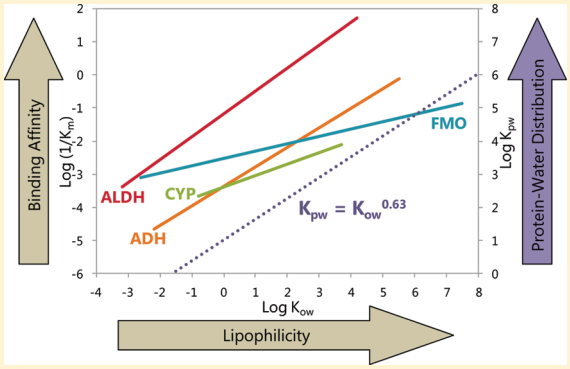
Alessandra Pirovano,*[,†] Mark A. J. Huijbregts,[†] Ad M. J. Ragas,[†,‡] and A. Jan Hendriks[†]

[†]Institute for Wetland and Water Research, Department of Environmental Science, Radboud University Nijmegen, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands.

[‡]School of Science, Open University, Heerlen, The Netherlands.

Supporting Information

**ABSTRACT:** In bioaccumulation models, biotransformation is one of the processes decreasing the concentration of chemicals in an organism. In order to be metabolized, a compound needs to bind to an enzyme. In this study, we derived relationships between binding affinity and lipophilicity, expressed as Log ($1/K_m$) and Log $K_{ow}$, respectively. We focused on oxidations in mammals catalyzed by alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), flavin-containing monooxygenase (FMO), and cytochrome P450 (CYP) enzymes. For all regressions, $1/K_m$ increased with compound $K_{ow}$, which can be understood from the tendency to biotransform lipophilic compounds into more polar, thus more easily excretable metabolites. Lipophilicity 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 0.63, typically noted in protein-water distribution (Log $K_{pw}$) and Log $K_{ow}$ regressions. A reduced slope (0.2−0.3) was found for FMO: this may be due to a different reaction mechanism involving a nucleophilic attack. The general patterns of metabolism were mechanistically interpreted in terms of partitioning theory. Information on the overall principles determining biotransformation may be helpful in predicting metabolic rates.




## INTRODUCTION

The EU REACH (registration, evaluation, authorization and restriction of chemicals) legislation[1] 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.[2]

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, for example, excretion via water, egestion via faeces, and growth dilution.[3] 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. First, the chemical needs to reach the enzyme and bind with it; second, 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.[4] The $K_m$ value is the substrate concentration at half the maximum rate, that is, at $V_{max}/2$, and is independent of the enzyme concentration.[5] The inverse of the Michaelis constant, that is, $1/K_m$, 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 oxidized by the microsomal cytochrome P450 (CYP).[6,7] The affinity, represented by $1/K_m$, was shown to be mainly related to the lipophilicity of the compound (see reviews in refs 4 and 8), although other factors might also be important, such as ionic interactions and hydrogen bonding properties.[9] However, these models focused on single CYP isoenzymes and small data sets, 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.[10] Nevertheless, the contribution of other enzymes to the oxidative metabolism of xenobiotics is significant as well.[11] Despite their importance,

Received: December 15, 2011
Revised: April 12, 2012
Accepted: April 12, 2012
Published: April 12, 2012





dx.doi.org/10.1021/es204506g | Environ. Sci. Technol. 2012, 46, 5168−5174

QSARs for non-CYP enzymes have not been developed. We hypothesized that the lipophilicity-binding regressions found for small data sets 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 catalyzed by alcohol dehydrogenase (ADH), aldehyde dehydrogenase (ALDH), flavin-containing monooxygenase (FMO), and CYP enzymes, in order to find generic patterns of metabolism across enzymes.

## MATERIALS AND METHODS

**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[12] (Braunschweig enzyme database, http://www.brenda-enzymes.org). $K_m$ values for FMO were taken from ref 13 and references contained therein. We also collected $K_m$ values for cytochrome P450 (CYP) from reviews.[6,14,15] 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 $\mu$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, nonrecombinant, 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[16] and CAS (Chemical Abstract Service) numbers were obtained for each compound from the ChemSpider Web site (http://www.chemspider.com/). The octanol–water partitioning coefficients ($K_{ow}$) were taken from the KOWWIN v 1.67, a program of EPI Suite available at the Web site of the U.S. Environmental Protection Agency (EPA) (http://www.epa.gov/opptintr/exposure/pubs/episuite.htm). Experimental $K_{ow}$ values, when available, were preferred over estimated ones. As the data sets included a number of compounds that would be ionized at physiological pH (7.4), we obtained Log $D_{7.4}$ values from ChemSpider, which are calculated using the software ACD Laboratories Log $D$ (Advanced Chemistry Development ACD/Laboratories Research, Toronto, Canada). The distribution coefficient $D_{7.4}$ represents the partitioning coefficient corrected for ionization of the chemical at pH 7.4.

Each compound was assigned to relevant chemical classes using the ECOSAR program v 1.0 present in EPI Suite. ECOSAR recognizes 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).[17]

The data collected ($K_m$ and $K_{ow}$) can be found in the Supporting Information (SI) (Table S1), with the references to the original papers.

**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 catalyze the same reaction but are characterized 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 characterized 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.

**Data Analysis.** Linear regression analysis was performed using the ordinary least squares (OLS) method. Among all data sets built with the different combinations of species/isoenzymes, we included in the analysis only those containing at least six compounds. For each data set, the QSAR equations were developed in the form:

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

We reported the slope (a) and the intercept (b) with their standard errors. The quality of the regression was characterized by the number of compounds used in the model ($n$), coefficient of determination ($r^2$), standard error for the estimated parameter 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 ionization in enzyme binding, we also developed the general regressions for the four enzyme families using Log $D_{7.4}$ values instead of 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 $p_{ancova}$ 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 metabolized by FMO and CYP for which we expected a similar behavior. 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 the SI (Table S2), together with their original ECOSAR classes and their general structure. The ECOSAR software does not separately categorize reactive chemicals such as OPs and CMs.[18] Therefore, we manually classified them and made a separate regression for OP pesticides, the only group with more than six 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, nonionizable, and nonreactive,[18] thus including diverse chemicals.

## RESULTS

All regressions made for each combination of isoenzyme and/or species are reported in Tables 1–4, 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 three letters. SI Tables S3–S6

**Table 1. Relationships between Log $K_{ow}$ and Log $(1/K_m)$ for ADH[a]**

| name | slope (±SE) | intercept (±SE) | $n$ | $r^2$ | SE | $p^b$ | $p_{ancova}{}^c$ |
|---|---|---|---|---|---|---|---|
| Regression Made Merging All Species (Mammals) and All Isoenzymes | | | | | | | |
| ADHgen | 0.59(±0.09) | −3.36(±0.18) | 34 | 0.56 | 0.82 | <0.01 | / |
| Regressions Made for the Separate Species (Mammals) and the Separate Isoenzymes | | | | | | | |
| ADH1_hor | 0.40(±0.11) | −3.08(±0.24) | 20 | 0.45 | 0.72 | <0.01 | 0.96 |
| ADH1_hum | 0.58(±0.12) | −3.01(±0.23) | 24 | 0.50 | 0.85 | <0.01 | 0.13 |
| ADH2_hum | 0.67(±0.19) | −3.58(±0.41) | 18 | 0.43 | 1.43 | <0.01 | 0.70 |
| ADH3_hum | 0.54(±0.25) | −4.38(±0.58) | 7 | 0.48 | 0.72 | *0.09* | *<0.01* |
| ADH1_rat | 0.62(±0.19) | −3.11(±0.32) | 13 | 0.50 | 0.84 | 0.01 | 0.28 |
| ADH3_rat | 1.18(±0.32) | −6.57(±0.75) | 6 | 0.77 | 0.82 | 0.02 | *<0.01* |
| Regression Made Merging All Species (Mammals) and All Isoenzymes, Using Log $D_{7.4}$ Values | | | | | | | |
| ADHgen ioniz | 0.60(±0.10) | −3.30(±0.18) | 34 | 0.52 | 0.85 | <0.01 | / |

[a]The $K_m$ values were expressed as $\mu$M. [b]The italic value indicates non significant regression ($p > 0.05$). [c]The italic values indicate regressions significantly different from ADHgen ($p_{ancova} < 0.05$).

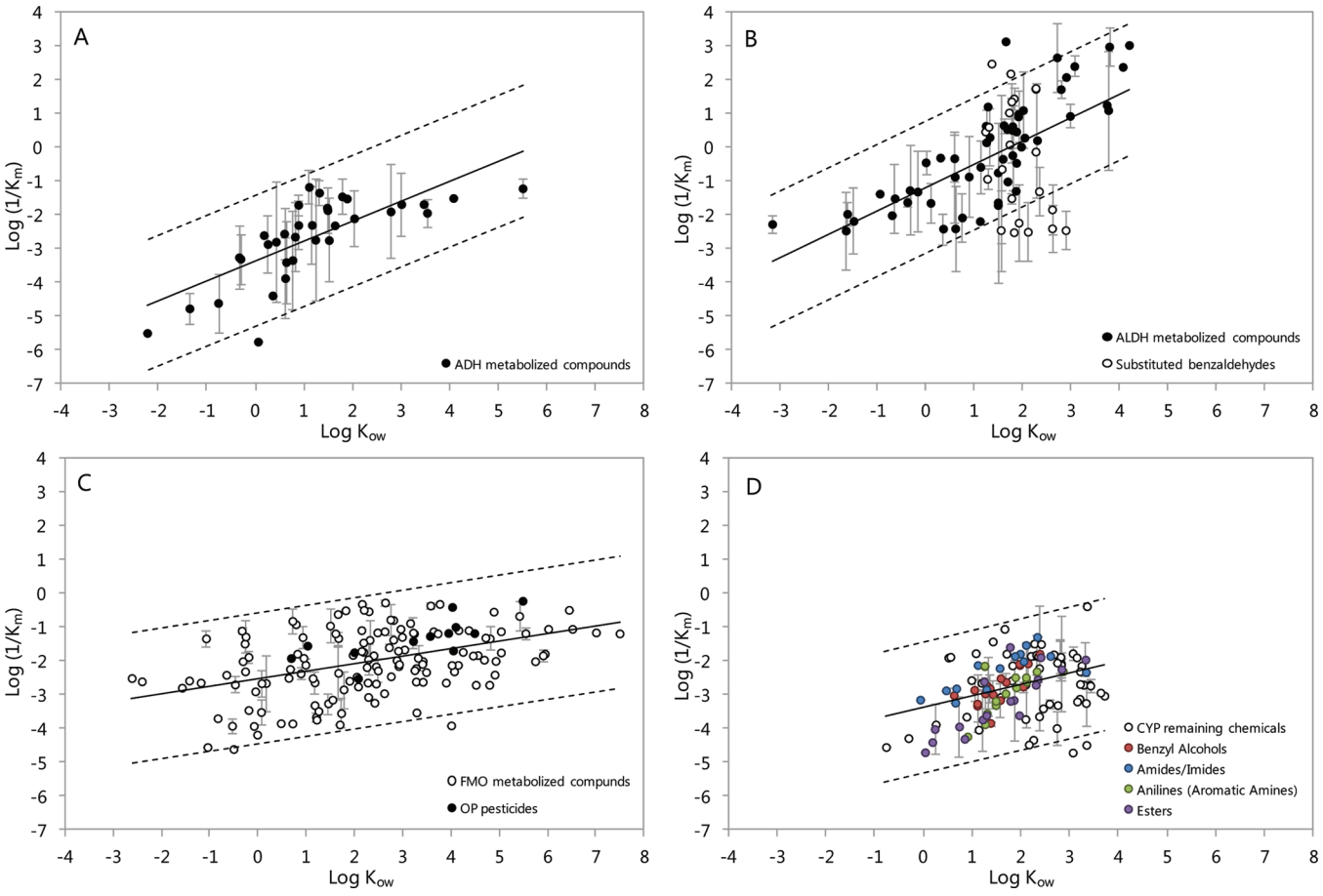



**Figure 1.** Relationships between Log $(1/K_m)$ and Log $K_{ow}$ in mammals for compounds metabolized by (A) ADH; (B) ALDH; (C) FMO; (D) CYP. Regressions (solid lines) and 95% confidence intervals (dashed lines). Laboratory measurements (dots): Log transformed geometrical mean of $1/K_m$ [$\mu M^{-1}$]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

provide 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.

**ADH.** We developed seven equations for ADH, which are reported in Table 1. The slope of the general regression ADHgen (Figure 1a) was 0.6, and the observed $K_m$ data were between 10 and $10^6$ $\mu$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 $p_{ancova} < 0.05$, the two regressions for ADH3 were statistically different from the general one; in particular, the intercepts were smaller. ADH data set 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.

**ALDH.** We initially built nine QSARs for ALDH, which are reported in Table 2. The general equation ALDHgen (Figure 1b) had a slope of 0.7, and the observed $K_m$ data were between $10^{-3}$ and $10^3$ $\mu$M. Among the specific regressions, the three 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 three equations for rat had $p_{ancova} < 0.05$. For ALDHgen, 11 out


dx.doi.org/10.1021/es204506g | Environ. Sci. Technol. 2012, 46, 5168−5174

**Table 2. Relationships between Log $K_{ow}$ and Log $(1/K_m)$ for ALDH, Together with three Additional General Regressions Leaving out the Possibly Influential Data: (I) Substituted Benzaldehydes; (II) Rat Data; (III) Rat Data As Well As Substituted Benzaldehydes[a]**

| name | slope (±SE) | intercept (±SE) | $n$ | $r^2$ | SE | $p$[b] | $p_{ancova}$[c] |
|---|---|---|---|---|---|---|---|
| Regression Made Merging All Species (Mammals) and All Isoenzymes | | | | | | | |
| ALDHgen | 0.69(±0.11) | −1.18(±0.22) | 77 | 0.33 | 1.33 | <0.01 | / |
| Regressions Made for the Separate Species (Mammals) and the Separate Isoenzymes | | | | | | | |
| ALDH1_hor | 0.99(±0.30) | −1.31(±0.38) | 10 | 0.57 | 1.00 | 0.01 | 0.84 |
| ALDH2_hor | 0.73(±0.35) | −0.43(±0.43) | 9 | 0.39 | 1.13 | *0.07* | 0.10 |
| ALDH1_hum | 0.82(±0.08) | −0.99(±0.17) | 28 | 0.80 | 0.73 | <0.01 | 0.19 |
| ALDH2_hum | 0.86(±0.13) | −0.73(±0.27) | 57 | 0.42 | 1.17 | <0.01 | *<0.01* |
| ALDH3_hum | 0.54(±0.17) | −1.18(±0.21) | 12 | 0.51 | 0.74 | 0.01 | 0.95 |
| ALDH1_rat | 0.18(±0.10) | −1.33(±0.17) | 32 | 0.10 | 0.73 | *0.08* | *<0.01* |
| ALDH2_rat | 0.10(±0.17) | −2.34(±0.26) | 22 | 0.02 | 1.00 | *0.55* | *<0.01* |
| ALDH3_rat | 0.56(±0.33) | −3.80(±0.74) | 8 | 0.32 | 0.45 | *0.14* | *<0.01* |
| I. Regression Made Merging All Species (Mammals) and All Isoenzymes, Excluding 22 Substituted Benzaldehydes | | | | | | | |
| | 0.81(±0.09) | −1.15(±0.17) | 55 | 0.63 | 0.96 | <0.01 | / |
| II. Regression Made Merging Horse and Human Data, And All Isoenzymes | | | | | | | |
| | 0.83(±0.10) | −0.84(±0.20) | 63 | 0.53 | 1.05 | <0.01 | / |
| III. Regression Made Merging Horse and Human Data, And All Isoenzymes, Excluding Substituted Benzaldehydes | | | | | | | |
| | 0.83(±0.09) | −0.92(±0.19) | 50 | 0.63 | 0.96 | <0.01 | / |
| Regression Made Merging All Species (Mammals) and All Isoenzymes, Using Log $D_{7.4}$ Values | | | | | | | |
| ALDHgen ioniz | 0.61(±0.12) | −1.00(±0.23) | 77 | 0.26 | 1.4 | <0.01 | / |

[a]The $K_m$ values were expressed as $\mu$M. [b]The italic values indicate non significant regressions ($p > 0.05$). [c]The italic values indicate regressions significantly different from ALDHgen ($p_{ancova} < 0.05$).

**Table 3. Relationships between Log $K_{ow}$ and Log $(1/K_m)$ for FMO, Together with an Additional Regression Developed Including Organophosphorous (OP) Pesticides only[a]**

| name | slope (±SE) | intercept (±SE) | $n$ | $r^2$ | SE | $p$ | $p_{ancova}$ |
|---|---|---|---|---|---|---|---|
| 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 Made for the Separate Species (Mammals) | | | | | | | |
| 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 Made for OP Pesticides, Merging All Species (Mammals) and All Isoenzymes | | | | | | | |
| | 0.32(±0.09) | −2.34(±0.33) | 12 | 0.54 | 0.45 | 0.01 | / |
| Regression Made Merging All Species (Mammals) and All Isoenzymes, Using Log $D_{7.4}$ Values | | | | | | | |
| FMOgen ioniz | 0.29(±0.04) | −2.43(±0.09) | 148[b] | 0.31 | 0.82 | <0.01 | / |

[a]The $K_m$ values were expressed as $\mu$M. [b]The Log $D_{7.4}$ value of one compound (2-aminoazulene) was not available.

of the total 77 compounds had observed $K_m$ values that were two orders of magnitude larger or smaller than expected from the regression. Nine of these outliers were substituted benzaldehydes. The ALDH data set contained 22 substituted benzaldehydes, which are represented by white dots in Figure 1b and listed in the SI (Table S7), together with their general structures.

We developed three additional general regressions leaving out the possibly influential data: (I) substituted benzaldehydes; (II) rat data; (III) rat data as well as substituted benzaldehydes. The three regressions (Table 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 rata data and substituted benzaldehydes were removed from the data set. 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 metabolized compounds; (2) including only rat data and excluding substituted benzaldehydes. The results are reported in the SI, Table S8 and Figure S1. 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 ($r^2 = 0.28$).

**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 three groupings (Table 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 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 1c), albeit with a shallow slope of 0.3 (Table 3).

**CYP.** For CYP, we first built five QSARs using all data (Table 4). The general equation CYPgen, which is illustrated in Figure 1d, had a slope of 0.3; the observed $K_m$ data were between 1 and $10^5$ $\mu$M. Among the separate regressions developed for the ECOSAR classes, poor correlation was found for the group of diverse chemicals, remaining chemicals, with $r^2 < 0.1$


dx.doi.org/10.1021/es204506g | Environ. Sci. Technol. 2012, 46, 5168−5174

Table 4. Relationships between Log $K_{ow}$ and Log $(1/K_m)$ for CYP, Together with Five Additional General Regressions for Separate ECOSAR classes: (I) Anilines (Aromatic Amines); (II) Benzyl Alcohols; (III) Esters; (IV) Amides/Imides; (V) Remaining Chemicals[a]

| name | slope (±SE) | intercept (±SE) | $n$ | $r^2$ | SE | $p^b$ | $p_{ancova}$ |
|---|---|---|---|---|---|---|---|
| Regression Made Merging All Species (Mammals) and All Isoenzymes | | | | | | | |
| CYPgen | 0.34(±0.08) | −3.38(±0.17) | 121 | 0.13 | 0.82 | <0.01 | / |
| Regressions Made for the Separate Species (Mammals) and the Separate Isoenzymes | | | | | | | |
| CYP1A1_rat | 0.52(±0.17) | −3.63(±0.32) | 23 | 0.30 | 0.54 | 0.01 | 0.75 |
| CYP2B1_rat | 0.08(±0.21) | −2.55(±0.48) | 39 | 0.00 | 1.02 | *0.70* | 0.09 |
| CYP2B4_rab | 0.24(±0.12) | −3.39(±0.27) | 47 | 0.08 | 0.76 | *0.05* | 0.12 |
| CYP2E1_rab | 0.78(±0.10) | −4.00(±0.16) | 36 | 0.65 | 0.51 | <0.01 | 0.94 |
| I. Regression Made for Anilines (Aromatic Amines), Merging All Species (Mammals) and All Isoenzymes | | | | | | | |
| | 0.77(±0.26) | −4.19(±0.46) | 17 | 0.37 | 0.51 | 0.01 | / |
| II. Regression Made for Benzyl Alcohols, Merging All Species (Mammals) and All Isoenzymes | | | | | | | |
| | 0.84(±0.20) | −4.03(±0.32) | 17 | 0.54 | 0.37 | <0.01 | / |
| III. Regression Made for Esters, Merging All Species (Mammals) and All Isoenzymes | | | | | | | |
| | 0.84(±0.14) | −4.48(±0.26) | 17 | 0.70 | 0.54 | <0.01 | / |
| IV. Regression Made for Amides/Imides, Merging All Species (Mammals) and All Isoenzymes | | | | | | | |
| | 0.48(±0.13) | −3.03(±0.23) | 14 | 0.54 | 0.43 | 0.01 | / |
| V. Regression Made for the Remaining Chemicals, Merging All Species (Mammals) and All Isoenzymes | | | | | | | |
| | 0.16(±0.13) | −3.02(±0.33) | 56 | 0.03 | 0.99 | *0.22* | / |
| Regression Made Merging All Species (Mammals) and All Isoenzymes, Using Log $D_{7.4}$ Values | | | | | | | |
| CYPgen ioniz | 0.25(±0.07) | −3.20(±0.15) | 121 | 0.10 | 0.83 | <0.01 | / |

[a]The $K_m$ values were expressed as $\mu$M. [b]The italic values indicate non significant regressions ($p > 0.05$).

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.

**Ionization.** 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 the SI (Table S9 and Figure S2). The 54% of the compounds in FMO data set 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 ionization improved the results only for FMO, although the correlation was still weak with a slope of 0.3 and $r^2 = 0.31$.

## DISCUSSION

**Regressions.** The QSAR models presented in this paper were developed for a well-defined end point ($K_m$), using an unambiguous algorithm that can be mechanistically interpreted, as recommended by OECD guidelines.[19] 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.[20] The lipophilicity parameter Log $K_{ow}$ combines these two properties.[21] 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.[20] The binding affinity increased with the compound $K_{ow}$ for four oxidizing enzymes tested in vitro in mammals (Tables 1−4), that is, 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, for example, 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.[6] 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).[22] Consequently, the input data are subject to variation, implying uncertainty in the regressions.

The data sets 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 S3−S6 in the SI. 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 data set. Furthermore, it is also recommended to check if the chemical belongs to one of the ECOSAR classes present in the data set.

We developed 24 QSARs, grouping the data according to two criteria: merging all species and all isoenzymes (four general regressions, one for each enzyme group), and separating each combination of a species and isoenzyme. In most cases, the four 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 three equations for ALDH in rat. In a previous study on ADH kinetics,[23] 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.[24] Regarding the regressions for ALDH in rat, Log $K_{ow}$ and 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 Log $D_{7.4}$ as descriptor, which represents the lipophilicity corrected for ionization of the chemical. The influence of ionization to binding affinity was relevant only for

compounds metabolized by FMO, for which the correlation with binding affinity increased, though slightly ($r^2$ = 0.31 and slope = 0.3). Therefore, the inclusion of Log $D_{7.4}$ did not contribute to improve the results significantly.

**Additional Regressions.** We developed nine 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 Log $K_{ow}$. These compounds had similar Log $K_{ow}$ values, ranging from 1.22 to 2.88, whereas their Log $(1/K_m)$ values covered five orders of magnitude, between −2.51 and 2.49. In the work of Klyosov,[25] the kinetics of ALDH toward 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 data sets contained mainly specific chemical classes, that is, 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 data sets. In Hansch's review on CYP,[6] 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 SI Table S10 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, seven gave acceptable regressions ($n > 6$, $p < 0.05$, underlined in SI Table S10). In the work of Lewis and Dickins,[26] 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.[9]

**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 S3, S4, and S6 in the SI) covering the value of 0.63, which is the typical slope correlating protein-water distribution (Log $K_{pw}$) and Log $K_{ow}$.[20] The value of 0.63 is in accordance with the slopes observed in other Log $K_{ow}$−Log$K_{pw}$ relationships, for example, 0.57 (for chemicals with Log $K_{ow}$ ranging from 2.0 to 5.1),[27] and about 0.7.[28] 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 $K_{ow}$.[20] 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 catalyze. ADH accepts a wide variety of substrates including exogenous primary and secondary alcohols, and oxidizes them to aldehydes and ketones, respectively. ALDH metabolizes endogenous and exogenous aldehydes to carboxylic alcohols (hydroxylation).[10] FMO catalyzes oxygenation of soft nucleophiles, that is, compounds with functional groups bearing a polarizable, electron-rich center, usually a heteroatom (such as nitrogen, sulfur and phosphorus) in organic compounds.[29] The poor correlation found for FMO could be attributed to its catalytic cycle, which is different with respect to the other enzymes.[13] 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.[30] In an extensive study on CYP3A4,[31] 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, whereas for N-, S-oxidation (mediated by FMO) a poor correlation was found between $K_m$ and $K_{ow}$.

**Application.** The regressions obtained in the present study relate the enzyme binding with 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 nonsaturating substrate concentration, the ratio between $V_{max}$ and $K_m$ provides an estimation of the intrinsic clearance ($CL_{int}$).[32,33] This parameter, which is a measure of enzyme activity toward a compound, can be extrapolated to equivalent whole-body metabolic rate.[34] Yet, in order to apply these regressions to predict whole-body metabolic rates, improvements are needed at various points. First, 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.

## ■ ASSOCIATED CONTENT

### Supporting Information

Original data, 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. Additional tables listing substituted benzaldehydes and DTC, OP, and CM pesticides, with their general chemical structure. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION


### Corresponding Author

*E-mail: a.pirovano@science.ru.nl.


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

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

## REFERENCES


(1) *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); L396/1*; Off. J. Eur. Union **2006**.

(2) Russom, C. L.; Breton, R. L.; Walker, J. D.; Bradbury, S. P. An overview of the use of quantitative structure-activity relationships for ranking and prioritizing large chemical inventories for environmental risk assessments. *Environ. Toxicol. Chem.* **2003**, *22* (8), 1810−1821.

(3) Hendriks, A. J.; van der Linde, A.; Cornelissen, G.; Sijm, D. T. H. M. The power of size. 1. Rate constants and equilibrium ratios for accumulation of organic substances related to octanol-water partition ratio and species weight. *Environ. Toxicol. Chem.* **2001**, *20* (7), 1399−1420.

(4) Testa, B.; Crivori, P.; Reist, M.; Carrupt, P.-A. The influence of lipophilicity on the pharmacokinetic behavior of drugs: Concepts and examples. *Perspect. Drug Discovery Des.* **2000**, *19* (1), 179−211.

(5) Sijm, D. T. H. M.; Rikken, M. G. J.; Rorije, E.; Traas, T. P.; McLachlan, M. S.; Peijnenburg, W. J. G. M. Transport, accumulation and transformation processes. In *Risk Assessment of Chemicals*, 2nd ed.; C. van Leeuwen, Vermeire, T., Eds.; Springer: Netherlands, 2007; pp 73−158.

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

(7) Lewis, D. F. V. Quantitative structure−activity relationships (QSARs) within the cytochrome P450 system: QSARs describing substrate binding, inhibition and induction of P450s. *Inflammopharmacology* **2003**, *11* (1), 43−73.

(8) Long, A.; Walker, J. D. Quantitative structure-activity relationships for predicting metabolism and modeling cytochrome P450 enzyme activities. *Environ. Toxicol. Chem.* **2003**, *22* (8), 1894−1899.

(9) Lewis, D. F. V.; Jacobs, M. N.; Dickins, M. Compound lipophilicity for substrate binding to human P450s in drug metabolism. *Drug Discov. Today* **2004**, *9* (12), 530−537.

(10) Klaassen, C. D. *Casarett and Doull's Toxicology: The Basic Science of Poisons*, 7th ed.; McGraw-Hills, 2008.

(11) Strolin Benedetti, M.; Whomsley, R.; Baltes, E. Involvement of enzymes other than CYPs in the oxidative metabolism of xenobiotics. *Expert Opin. Drug Metab. Toxicol.* **2006**, *2* (6), 895−921.

(12) Scheer, M.; Grote, A.; Chang, A.; Schomburg, I.; Munaretto, C.; Rother, M.; Söhngen, C.; Stelzer, M.; Thiele, J.; Schomburg, D. BRENDA, the enzyme information system in 2011. *Nucleic Acids Res.* **2011**, *39* (suppl 1), D670−D676.

(13) Krueger, S. K.; Williams, D. E. Mammalian flavin-containing monooxygenases: Structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol. Therapeut.* **2005**, *106* (3), 357−387.

(14) Hansch, C.; Zhang, L. Quantitative structure-activity relationships of cytochrome P-450. *Drug Metab. Rev.* **1993**, *25* (1−2), 1−48.

(15) Lewis, D. F. V. Frontier orbitals in chemical and biological activity: Quantitative relationships and mechanistic implication. *Drug Metab. Rev.* **1999**, *31* (3), 755−816.

(16) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comp. Sci.* **1988**, *28* (1), 31−36.

(17) Madden, J. C.; Enoch, S. J.; Hewitt, M.; Cronin, M. T. D. Pharmaceuticals in the environment: Good practice in predicting acute ecotoxicological effects. *Toxicol. Lett.* **2009**, *185* (2), 85−101.

(18) Zvinavashe, E.; Murk, A. J.; Rietjens, I. M. C. M. On the number of EINECS compounds that can be covered by (Q)SAR models for acute toxicity. *Toxicol. Lett.* **2009**, *184* (1), 67−72.

(19) *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*; OECD (Organisation for Economic Co-operation and Development): Paris, 2006.

(20) Hendriks, A. J.; Traas, T. P.; Huijbregts, M. A. J. Critical body residues linked to octanol−water partitioning, organism composition, and LC50 QSARs: Meta-analysis and model. *Environ. Sci. Technol.* **2005**, *39* (9), 3226−3236.

(21) Lewis, D. F. V. Structural characteristics of human P450s involved in drug metabolism: QSARs and lipophilicity profiles. *Toxicology* **2000**, *144* (1−3), 197−203.

(22) Cronin, M. T. D.; Schultz, T. W. Pitfalls in QSAR. *J. Mol. Struct.: THEOCHEM* **2003**, *622* (1−2), 39−51.

(23) Deetz, J. S.; Luehr, C. A.; Vallee, B. L. Human liver alcohol dehydrogenase isozymes: Reduction of aldehydes and ketones. *Biochemistry* **1984**, *23* (26), 6822−6828.

(24) Holmes, R. Alcohol dehydrogenases: Gene multiplicity and differential functions of five classes of isozymes. *Drug Alcohol Rev.* **1993**, *12* (1), 99−110.

(25) Klyosov, A. A. Kinetics and specificity of human liver aldehyde dehydrogenases toward aliphatic, aromatic, and fused polycyclic aldehydes. *Biochemistry* **1996**, *35* (14), 4457−4467.

(26) Lewis, D. F. V.; Dickins, M. Baseline lipophilicity relationships in human cytochromes P450 associated with drug metabolism. *Drug Metab. Rev.* **2003**, *35* (1), 1−18.

(27) deBruyn, A. M. H.; Gobas, F. A. P. C. The sorptive capacity of animal protein. *Environ. Toxicol. Chem.* **2007**, *26* (9), 1803−1808.

(28) Schwarzenbach, R. P.; Gschwend, P. M.; Imboden, D. M. Sorption II: Partitioning to living media - Bioaccumulation and baseline toxicity. In *Environmental Organic Chemistry*, 2nd ed.; Wiley-Interscience: New York, 2002; pp 331−386.

(29) Ziegler, D. M. Flavin-containing monooxygenases: Enzymes adapted for multisubstrate specificity. *Trends Pharmacol. Sci.* **1990**, *11* (8), 321−324.

(30) Guengerich, F. P. Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem. Res. Toxicol.* **2001**, *14* (6), 611−650.

(31) Bu, H. Z. A literature review of enzyme kinetic parameters for CYP3A4-mediated metabolic reactions of 113 drugs in human liver microsomes: Structure- kinetics relationship assessment. *Curr. Drug Metab.* **2006**, *7* (3), 231−249.

(32) Lipscomb, J. C.; Poet, T. S. In vitro measurements of metabolism for application in pharmacokinetic modeling. *Pharmacol. Therapeut.* **2008**, *118* (1), 82−103.

(33) Weisbrod, A. V.; Sahi, J.; Segner, H.; James, M. O.; Nichols, J.; Schultz, I.; Erhardt, S.; Cowan-Ellsberry, C.; Bonnell, M.; Hoeger, B. The state of in vitro science for use in bioaccumulation assessments for fish. *Environ. Toxicol. Chem.* **2009**, *28* (1), 86−96.

(34) Nichols, J. W.; Schultz, I. R.; Fitzsimmons, P. N. 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. *Aquat. Toxicol.* **2006**, *78* (1), 74−90.