Environmental Chemistry

ESTIMATION OF THE SOIL–WATER PARTITION COEFFICIENT NORMALIZED TO ORGANIC CARBON FOR IONIZABLE ORGANIC CHEMICALS

ANTONIO FRANCO* and STEFAN TRAPP
Department of Environmental Engineering, Technical University of Denmark, Miljøvej, Building 113, DK-2800 Kgs. Lyngby, Denmark

(Received 6 November 2007; Accepted 10 March 2008)

Abstract—The sorption of organic electrolytes to soil was investigated. A dataset consisting of 164 electrolytes, composed of 93 acids, 65 bases, and six amphoters, was collected from literature and databases. The partition coefficient log $K_{oc}$ of the neutral molecule and the dissociation constant $pK_a$ were calculated by the software ACD/Labs®. The Henderson-Hasselbalch equation was applied to calculate dissociation. Regressions were developed to predict separately for the neutral and the ionic molecule species the distribution coefficient ($K_d$) normalized to organic carbon ($K_{oc}$) from log $K_{ow}$ and $pK_a$. The log $K_{oc}$ of strong acids ($pK_a < 4$) was not correlated to these parameters. The regressions derived for weak acids and bases (undissociated at environmental pH) were similar. The highest sorption was found for strong bases ($pK_a > 7.5$), probably due to electrical interactions. Nonetheless, their log $K_{oc}$ was highly correlated to log $K_{ow}$. For bases, a nonlinear regression was developed, too. The new regression equations are applicable in the whole $pK_a$ range of acids, bases, and amphoters and are useful in particular for relatively strong bases and amphoters, for which no predictive methods specifically have been developed so far.

Keywords—Acids Bases Electrolytes Soil Sorption

INTRODUCTION

The distribution coefficient ($K_d$) describes the equilibrium concentration ratio of chemicals between soil, sediment, or particles and water. Therefore, it is a key parameter for mobility and fate of environmental chemicals. Many transport processes, as well as bioavailability and degradation rates, are directly related to this parameter. It has been recognized that, for organic chemicals, the $K_d$ is highly correlated to the fraction of organic carbon in soils or sediments [1], $K_d = f_{oc} \cdot K_{oc}$ where $f_{oc}$ is the fraction of organic carbon (g/g dry wt based), and $K_{oc}$ is the soil–water partition coefficient normalized to organic carbon (L/kg). In literature the same notation $K_{oc}$ sometimes is used for the real water–organic carbon partition coefficient, sometimes for the soil–water partition coefficient normalized to organic carbon. The two quantities are identical only if sorption occurs exclusively to the soil organic fraction. In the present study we adopted the latter notation.

The $K_{oc}$ is correlated to the octanol–water partition coefficient, $K_{ow}$; this is particularly true for hydrophobic compounds. Based on the work of Sabljic et al. [2], 19 quantitative structure-activity relationships for a variety of chemical classes are listed in the European Union technical guidance document for risk assessment of chemicals [3] in the form

$$\log K_{oc} = a \cdot \log K_{ow} + b$$  (1)

Possible processes behind the sorption of organic chemicals to soil and sediment are ion bonding or ligand exchange, chemisorption (formation of a bond, usually covalent, with the soil molecular structure), ion–dipole and dipole–dipole interactions, charge transfer, hydrogen bonding, and hydrophobic bonding (Van der Waals forces) [4]. The most chemically active component of the soil is the colloidal fraction, which consists of organic matter and inorganic clay minerals [5]. Both components display a negative electrical charge at the surface. The effect of this charge can be measured by the cationic exchange capacity, which on average is 50 meq/100 g for clays and 290 meq/100 g for humic acids [6]. Electrical forces involving charge transfer (~40 kJ/mol) are stronger than hydrophobic bonding (~4 kJ/mol) [4] so that they dominate when present. Thus, a different degree of sorption of anions, cations, and neutral molecules can be expected, with cations showing the highest potential for sorption, due to electrical attraction. Although lipophilic interactions are weaker than the others mentioned, they are the most important for the majority of organic chemicals.

In the technical guidance document [3], two regressions of the type of Equation 1 are listed that are applicable for organic acids, with $a = 0.47$ and $b = 0.50$ ($r^2$ is 0.72) for the first equation derived for 36 alcohols and organic acids, and $a = 0.60$ and $b = 0.32$ ($r^2$ is 0.75) for the second regression derived for 23 organic acids [2]. The log $K_{ow}$ used for the regression was a measured value for most compounds; when this value was not available it was calculated from the structure. In both cases it refers to the log $K_{ow}$ of the neutral molecule. Some quantitative structure-activity relationships include in their applicability domain chemical classes that comprise weak bases, such as triazines, anilines, and amides. However, no regression specifically applicable to bases, amphoters, or other electrolytes is listed.

Bintein and Devillers [7] suggested an empirical regression using $K_{ow}$ and $pK_a$ as molecule descriptors and pH and $f_{oc}$ to describe the sorbent properties for the estimation of the $K_d$ of both ionized and nonionized chemicals

$$\log K_d = 0.93 \cdot \log K_{ow} + 1.09 \cdot \log f_{oc} + 0.32 \cdot CF_a$$
$$- 0.55 \cdot CF_b + 0.25$$  (2)

where the correction factors $CF_a$ and $CF_b$ quantify the variation of dissociated acids and bases in the system.
According to Bintein and Devillers [7], for acids, the soil pH should be entered into the equation for the correction factor $CF_a$. The pH required for the correction factor $CF_a$ is the pH at the surface of the soil colloids and is lowered 2 pH units below the pH of the bulk soil solution. Even though the coefficient of determination of this equation was very high ($r^2 = 0.93$), a shortcoming of the regression, regarding electrolytes, is their insufficient representation in the test set (nine acids and three bases out of 87 chemicals).

No rigorously tested method is available to predict the $K_d$ or $K_{OC}$ of the majority of electrolytes. The present study provides regressions to predict the $K_d$ normalized to organic carbon ($K_{OC}$) for organic acids, bases, and amphoters from log $K_{OW}$ and $pK_a$. Taking into account dissociation, separate regression equations for the neutral and the ionic molecule fraction were derived and tested.

### MATERIALS AND METHODS

#### Data collection

Experimental values for $K_{OC}$ were selected from databases and literature sources (the complete dataset with references is available in the Supporting Information; http://dx.doi.org/10.1897/07-583.S1). The dataset comprises pesticides, industrial chemicals and pharmaceuticals. Experimental $K_{OC}$ (L/kg) are calculated dividing $K_d$ values by the fraction of organic carbon $f_{OC}$

$$K_{OC} = \frac{K_d}{f_{OC}}$$

The $K_{OC}$ defined by Equation 4 is not exactly equal to the distribution coefficient between organic carbon and water, unless sorption occurs exclusively onto the organic fraction of the soil or sediment. Databases often report an average $K_{OC}$ calculated from experimental $K_d$ measured in soils having different $f_{OC}$. If more than one $K_{OC}$ value was found for a compound, the average was used. In total, 164 values for dissociating compounds were collected: 93 acids, 65 bases, and six amphoters (Table 1). For neutral hydrophobic compounds, the variation of experimental $K_{OC}$ values of different soils is within one order of magnitude, whereas the variation in the $K_{OC}$ of polar and ionisable compounds can be up to two orders of magnitude [4]. Therefore we preferably collected averaged values from a representative variety of soils.

Because we aimed at treating the ionic and the neutral fraction separately, it was essential to define the octanol–water partition coefficient of the neutral molecule ($P_{\text{neutral}}$) and of the ion ($P_{\text{ion}}$). The overall $K_{OW}$ at a given pH is calculated from the neutral ($\phi_n$) and the ionic fraction ($\phi_{ion}$)

$$K_{OW} = \phi_n P_n + \phi_{ion} P_{ion}$$ (5)

Often, log $K_{OW}$ values published in literature are apparent log $K_{OW}$ values (log $D$) at the experimental pH. Also, the log $K_{OW}$ of the neutral molecule occasionally is difficult to determine (e.g., for strong acids or bases). Therefore we preferred software estimations for the log $K_{OW}$ of the neutral compound fraction. The software ACD/Labs® (ACD/L- Lab, ver 6.01, Advanced Chemistry Development, Toronto ON, Canada) was chosen based on the review of existing programs for the estimation of log $P_n$, provided by the European Union (REACH Implementation Project 3.3 Phase 2, final draft, available at http://ecb.jrc.it/reach/rip/). The program estimates log $P_n$ and the dissociation constant(s) $pK_a$, using a structure-fragment algorithm. The dissociation constant(s) is calculated at 25°C and zero ionic strength. We selected chemicals with acidic $pK_a < 12$ and basic $pK_a > 2$. Outside this $pK_a$ range (very weak acids and bases) the effect of electrical forces involving ions can be neglected because the compound almost completely is nondissociated. Figure 1 illustrates the chemical property space of the acids and bases considered. In case the ACD software failed to calculate the $pK_a$ (e.g., organotins), measured values were selected. For cationic groups completely dissociated in water, a default $pK_a$ of 12 was chosen. For completely dissociated bases ($pK_a > 12$), measured $K_{OW}$ values were taken as $P_{ion}$ and $P_n$ was calculated according to Trapp and Horobin [8]

$$\log P_{\text{ion}} = \log P_n - 3.5$$ (6)

Although log $P_{\text{ion}}$ could be calculated from the log $D$ at appropriate pH, the correlation between log $P_n$ and log $P_{\text{ion}}$ was too high to justify the introduction of a new variable in the regressions.

#### Statistical methods

Regressions were developed to predict log $K_{OC}$ from log $P_n$ and $pK_a$. Linear regressions were fitted by the least squares method using the data analysis package in Excel® (Microsoft, Redmond, WA, USA). Nonlinear regressions were fitted using the nonlinear least squares function in the software R for statistical computing (ver 2.5.1) [9]. A training set of 62 acids and 43 bases was used for fitting; randomly selected independent datasets of 31 acids and 22 bases were used to validate the regressions. The dataset for amphoters included six compounds.

The preciseness (the distance between predicted and experimental values) was evaluated by the root mean square deviation (RMSD) between predicted and experimental data

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}}$$ (7)

where $y_i$ is the predicted value, $x_i$ is the measured value, and $n$ is the number of events. The common variance of predictor and estimate was evaluated by the coefficient of determination, $r^2$. The significance of correlations was judged using tabulated values of the critical correlation coefficient for varying levels of significance $\alpha$ [10].

The accuracy of the prediction methods (the agreement between the estimates and real values) was evaluated by the slope
and the intercept of the regression between predicted and measured \(K_{oc}\) values.

**Calculation of dissociation**

**Monovalent acids and bases.** The speciation of monovalent ionic compounds was calculated according to the Henderson-Hasselbalch equation

\[
\phi_a = \frac{1}{1 + 10^{\text{pH}-pK_{a}}} \\
\phi_n = 1 - \phi_a
\]

where \(a = 1\) for acids and \(-1\) for bases.

**Amphoters.** Amphoters are defined as chemicals that can react as either an acid or a base. It is necessary to distinguish between substances that mostly are undissociated in the neutral form and substances whose net neutral fraction displays simultaneously positive and negative charge (zwitterions). Herein we define as amphoters the substances of the former type (\(pK_{a,\text{base}} < \text{pH} < pK_{a,\text{acid}}\)) and as ampholytes the substances that exist mostly as zwitterions within a certain pH range (\(pK_{a,\text{acid}} < \text{pH} < pK_{a,\text{base}}\)). Ampholytes are not treated in the present study. The Henderson-Hasselbalch equation for the speciation of an amphoter is

\[
\phi_a = \frac{1}{1 + 10^{\text{pH}-pK_{a}}} \\
\phi_n = \phi_a \cdot 10^{pK_{a}-\text{pH}} \\
\phi_c = \phi_n \cdot 10^{pK_{n}-\text{pH}}
\]

where \(\phi_a\) and \(\phi_n\) are the anionic and cationic fractions, respectively. A fraction of the neutral species may exist as zwitterion when the two dissociation constants are close. We neglected the minor zwitterionic fraction of amphoters because the basic \(pK_a\) was at least two units lower than the acidic one for the six amphoters considered.

**RESULTS**

**Correlation between log \(K_{oc}\) and \(pK_a\) in the dataset.**

In the plot of \(pK_a\) versus log \(K_{oc}\) (Fig. 2) no clear trend shows up. In fact, the linear correlation between \(pK_a\) and log \(K_{oc}\) is insignificant both for acids and bases (\(\alpha > 0.1\)). Acids with low \(pK_a\) are dissociated at environmental pH. With a few exceptions, the measured log \(K_{oc}\) of strong acids ranges from 10 to 300 L/kg (log \(K_{oc}\) between 1 and 2.5). Acids with high \(pK_a\) on average show a higher sorption than the ones with low \(pK_a\). Conversely, bases with low \(pK_a\) are neutral, and bases dissociate when they have a high \(pK_a\). On average, the sorption of bases increases with \(pK_a\) until it reaches approximately 8 (Fig. 2). For higher \(pK_a\), the log \(K_{oc}\) fluctuates between extreme values without clear relation to \(pK_a\).

The correlation between log \(K_{oc}\) and log \(K_{oc}\) of acids being neutral (\(pK_a > 7.5\)) or dissociated (\(pK_a < 4\)) at typical soil pH was investigated separately (Fig. 3). The plot of log \(P_{ow}\) (the log \(K_{oc}\) of the dissociated molecule species) versus log \(K_{oc}\) for anions (Fig. 3a) shows no significant relation (\(r^2 = 0.09, \alpha > 0.1\)). Contrarily, the correlation between log \(P_{ow}\) (the log \(K_{oc}\) of the neutral molecule species) versus log \(K_{oc}\) is strong (\(r^2 = 0.81, \alpha < 0.001\)) for the weak acids (Fig. 3c) that are neutral at soil pH.

The pattern is different for bases (Fig. 3b and d). The linear correlation between the log \(P_{ow}\) to the log \(K_{oc}\) is highly significant (\(r^2 = 0.72, \alpha < 0.001\)) for cations. A strong correlation (\(r^2 = 0.62, \alpha < 0.001\)) also was found between log \(P_{n}\) and log \(K_{oc}\) for neutral, weak bases. It can be observed that, at similar lipophilicity, the sorption of cations is stronger than the sorption of neutral bases, i.e., strong sorption of cations occurs even at relatively low values of log \(P_{ow}\). This makes it unlikely that the process behind lipophilic sorption. We speculate that the driving force behind the strong sorption of cations is electrical attraction. However, the good correlation between log \(K_{oc}\) and log \(P_{ow}\) suggests that lipophilicity, at least indirectly, has a strong impact on the sorption of cations.

**Formulation of regressions**

The sorption increases with log \(P_{ow}\) or log \(P_{on}\) in all four cases. The trend is highly significant for neutral molecules and cations, but insignificant for anions. It may be concluded that the sorption of weak electrolytes, which are neutral under environmental pH, is mostly lipophilic, because the log \(K_{oc}\) correlates well with the log \(P_{ow}\). Subsequently, their log \(K_{oc}\) can be predicted from the log \(P_{on}\) using the conventional approach

\[
\log K_{oc} = a \cdot \log P_{on} + b
\]

The sorption of anions generally is moderate but not negligible, even for very hydrophilic anions. The overall \(K_{oc}\) of dissociating chemicals results from the sum of the contributions of the neutral (\(K_{oc,n}\)) and the ionic fraction (\(K_{oc,i}\)).
Applying Equation 10 both for the neutral and the ionic molecule fraction gives

\[ K_{OC} = \phi_n \cdot K_{OC,n} + \phi_{ion} \cdot K_{OC,ion} \]  

(11)

The first term measures the lipophilic absorption of the neutral fraction, the second the contribution of the ionic fraction (which is probably the sum of lipophilic absorption and electrical adsorption). Because \( \log P_{\text{ion}} \) is related to \( \log P_n \) (Eqn. 6), it can be replaced by \( \log P_n \) when fitting Equation 12 to the experimental data by using a linear regression.

Results of the fits (Eqns. 13–21) are reported in Table 2 together with quality parameters. Two separate approaches were used for both acids and bases to best fit Equation 12 to experimental \( \log K_{OC} \) values. In the first approach, we fit separately the exponentials in the first and the second term of Equation 12 with a linear regression in the form of Equation 10, using the dataset of weak and strong acids (Eqns. 13 and 14 in Table 2) and bases (Eqns. 17 and 18 in Table 2). No dissociation for weak acids (pK_a > 7.5) and bases (pK_a > 3.5) and complete dissociation for strong acids (pK_a < 4) and bases (pK_a < 7.5) were assumed. The two regressions subsequently were applied for the entire training set (combined regression, Eqns. 15 and 19 in Table 2) to determine the pH that gave the best fit (pH_opt). The parameter pH_opt and its implication on the model domain are further explained in the Discussion section.

In the second approach, we directly fitted Equation 12 using the full training set to determine all parameters at once (direct regression, Eqns. 16 and 20 in Table 2).

For acids the two approaches gave very similar equations; pH_opt was 5.8 in both cases. For bases, the two approaches yielded different parameters. Equation 20 determined by the direct method (pH_opt is 6) fit slightly better the training dataset than the combined regression (pH_opt is 4). The better agreement was confirmed by validation. Figure 4 shows the predicted \( \log K_{OC} \) as a function of \( \log P_n \) and pK_a for acids (Eqn. 15) and for bases (Eqn. 20). Although \( \log K_{OC} \) always increases with \( \log P_n \), the effect of dissociation is more complex. Sorption increases with pK_a for both lipophilic acids and bases, but it decreases with pK_a for hydrophilic. This means that cations sorb more than the corresponding neutral molecule, but anions do less.

Furthermore, the similarity of the regressions found for weak acids and weak bases (Eqns. 13 and 17) suggested deriving a single regression for the neutral fraction of weak electrolytes (Eqn. 21 in Table 2) from all weak acids and bases in the training set. For comparison, the result for the regression suggested in the technical guidance document for the validation dataset also is shown (for acids only). The correlation coefficients of all regressions are highly significant (\( \alpha < 0.001 \)), except for Equation 14 (strong acids, not significant, \( \alpha > 0.1 \)).

**DISCUSSION**

**Goodness of fit**

**Acids.** Calculated versus experimental \( \log K_{OC} \) using Equation 15 are shown in Figure 5 (training and validation set).
Table 2. Regression Equations 13 to 21 for acids and bases calculated to fit Equation 12 (combined and direct approach) and regression suggested for weak acids by the European Union technical guidance document (TGD) [3]. The RMSD is the root mean square deviation. Log $P_n$ is the log $K_{OC}$ of the neutral molecule. In brackets: Result for the validation set. $\phi_n$: Neutral fraction; $\phi_{an}$: Ionic fraction.

<table>
<thead>
<tr>
<th>Type</th>
<th>Eqn.</th>
<th>Regression</th>
<th>$n$</th>
<th>RMSD</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak acids</td>
<td>13</td>
<td>$\log K_{OC} = 0.54 \cdot \log P_n + 1.11$</td>
<td>11</td>
<td>0.51</td>
<td>0.81</td>
</tr>
<tr>
<td>Strong acids</td>
<td>14</td>
<td>$\log K_{OC} = 0.11 \cdot \log P_n + 1.54$</td>
<td>29</td>
<td>0.54</td>
<td>0.04</td>
</tr>
<tr>
<td>Full training set acids</td>
<td>15</td>
<td>$\log K_{OC} = \log (\phi_{an} \cdot 10^{0.54 \cdot \log P_n + 1.11} + \phi_{an} \cdot 10^{0.11 \cdot \log P_n + 1.54})$</td>
<td>62</td>
<td>0.47</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>$\log K_{OC} = \log (\phi_{an} \cdot 10^{0.52 \cdot \log P_n + 1.27} + \phi_{an} \cdot 10^{0.47 \cdot \log P_n + 1.69})$</td>
<td>62</td>
<td>0.45</td>
<td>0.54</td>
</tr>
<tr>
<td>Full training set acids</td>
<td>17</td>
<td>$\log K_{OC} = 0.6 \cdot \log P_n + 0.32$</td>
<td>(31)</td>
<td>0.74</td>
<td>(0.57)</td>
</tr>
<tr>
<td>TGD regression (acids only)</td>
<td>18</td>
<td>$\log K_{OC} = 0.59 \cdot \log P_n + 1.34$</td>
<td>11</td>
<td>0.26</td>
<td>0.83</td>
</tr>
<tr>
<td>Strong bases</td>
<td>19</td>
<td>$\log K_{OC} = 0.47 \cdot \log P_n + 1.95$</td>
<td>13</td>
<td>0.59</td>
<td>0.75</td>
</tr>
<tr>
<td>Full training set bases</td>
<td>20</td>
<td>$\log K_{OC} = \log (\phi_{an} \cdot 10^{0.21 \cdot \log P_n + 1.24} + \phi_{an} \cdot 10^{0.22 \cdot \log P_n + 2.19})$</td>
<td>43</td>
<td>0.49</td>
<td>0.60</td>
</tr>
<tr>
<td>Full training set bases</td>
<td>21</td>
<td>$\log K_{OC} = 0.50 \cdot \log P_n + 1.13$</td>
<td>22</td>
<td>0.46</td>
<td>0.62</td>
</tr>
<tr>
<td>Weak electrolytes (neutral</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fraction)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The RMSD and $r^2$ are reported in Table 2. Although the precision (RMSD) is not very high, the equation gives acceptable estimates of experimental values for both the training and validation sets. The precision of the regressions is limited by the apparently random variability in the sorption of the anions, which is not correlated with log $P_n$ (also seen from the small slope on log $P_n$ for the anionic fraction in Eqns. 14 and 15, which is only 0.11). Although there is no statistical significance in the log $K_{OC}$–log $P_n$ relationship for anions and the anionic contribution could be replaced by a constant mean value, Equation 15 performed slightly better in the full form. This requires no additional variable.

The performance of the regression was compared to the equation suggested for organic acids in the TGD (Fig. 5e) using the validation set. Even though the explained variance ($r^2$) is high, the technical guidance document regression underestimates the sorption of moderate-to-strong acids and has an average error (RMSD) of 0.74. Also, slope and intercept of the regression between predicted and experimental log $K_{OC}$ differ considerably from the values 1 and 0, which would show perfect agreement. In fact, its applicability is restricted to weak acids [2].

Four acids are marked with triangles (for which the log $K_{OC}$ is underestimated by more than two log units) out of the typical variation of experimental measurements [4]. These are fluoroxypyr ($pK_a = 2.22$, log $P_n = 3.16$) in the training set, sintofen ($pK_a = 0.16$, log $P_n = 1.89$), chlorophacinone ($pK_a = 4.5$, log $P_n = 5.22$), and fenoxaprop-$p$-ethyl ($pK_a = 3.14$, log $P_n = 4.18$) in the validation set. Specific electrical affinity of certain functional groups may explain the high log $K_{OC}$ of fluoroxypyr and sintofen. For the last two compounds (with intermediate $pK_a$ and high log $P_n$), experimental $K_{OC}$ possibly were measured in acidic soil samples, where the lipophilic neutral molecule was nondissociated.

**Bases.** Calculated versus experimental log $K_{OC}$ using the direct regression (Eqn. 20) are shown in Figure 5b (training set) and Figure 5d (validation set). The RMSD and $r^2$ are reported in Table 2. Equation 20 fits accurately the experimental log $K_{OC}$ of the training set. This result is confirmed by validation. The precision is satisfactory, although it seems to decrease with increasing sorption. Two outliers are marked with triangles, representing acridine and benzo[f]quinoline, whose deviation from experimental $K_{OC}$ is between 1.4 and 1.8 log units. Even if the deviation is within the possible variation of two log units in experimental measurements for different sorbents [4], the log $K_{OC}$ is a poor descriptor of sorption for these two nitrogen heterocyclic aromatic bases because they strongly adsorb to clay [11]. Despite their relatively low $pK_a$ (5.6 and 5.15, respectively), these compounds can dissociate at the surface of clay minerals because in this region the pH can be 3 to 4 pH units below the bulk pH [12]. The capability to delocalize charges through the three aromatic
Amphoters. If the speciation of amphoters is calculated according to Equation 9, it is possible to sum the contribution of the sorption of the anionic, cationic, and neutral fraction using Equations 14, 24 (only the term for the cationic fraction), and 21, respectively, to estimate the log $K_{OC}$ of amphoters. The results were compared with measured log $K_{OC}$ values of six compounds (Fig. 5f). The few available data agree satisfactorily with the estimates (RMSD = 0.51) and the optimal


Fig. 6. Influence of log $K_{OC}$ calculated at pH 7 (log D) on the sorption of bases with $pK_a > 4$ (full dataset). The vertical line corresponds to $K_{OW} = 1$ (log $K_{OW} = 0$).

pH is 5. This shows that the regressions obtained for anions, cations, and the neutral fraction of weak electrolytes are of general validity.

**Weaknesses of approaches**

Existing regressions derived from a single parameter ($K_{OW}$) in the form of Equation 1 do not account for the different sorption behavior of the neutral molecule and the ion. In equations of this type, the $pK_a$ applicability range is not specified. By considering separately the contribution of the neutral and ionic species to sorption, the regressions become more accurate. As shown in Figure 3, log $P_a$ (in our notation, the log $K_{OW}$ of the ion) is not a good descriptor for the sorption of anions. This limits the precision of Equation 15, in particular for strong acids. Equation 20 is precise satisfactorily for strong bases because log $P_a$ (thus also log $P_{an}$) correlates well with the sorption of strong cations (Fig. 3). However, the linear relationship between log $K_{OC}$ and log $P_a$ (Eqn. 18) does not describe precisely the variability of the log $K_{OC}$ of strong cations (RMSD is 0.59). We therefore investigated further the effect of $K_{OW}$ and $pK_a$ on the sorption of strong bases.

**Development of the regression for bases**

As outlined in the Results section, the sorption of relatively hydrophilic cations is high and, even though the correlation to log $P_{an}$ is highly significant, it is unlikely that the driving force is lipophilic sorption. Figure 6 shows the relation between log $K_{OW}$ calculated at pH 7 (the real log $K_{OW}$ of the sum of neutral molecule species and cation at a given pH; often named log $D$) and log $K_{OC}$ for bases with $pK_a > 4$ (bases, for which the cation dominates the sorption). Generally, the sorption increases with log $K_{OW}$, but surprisingly, the correlation between log $K_{OW}$ and log $K_{OC}$ is stronger for the hydrophilic compounds (log $K_{OW} < 1$). An explanation may be that the log $K_{OW}$ has an impact on the transport of the cations to the sorption sites. Considering the path a molecule takes from the bulk soil pore water to reach the binding sites at the surface of organic colloids, we hypothesize that the transport of hydrophilic molecules from the bulk soil solution into the organic matter partly is limited by diffusion through the organic layer itself. The tendency of a molecule to penetrate into the organic matter is proportional to its $K_{OW}$ [13] and, for hydrophilic compounds, the diffusion process may be limiting the sorption (at least in time-limited experiments). In Figure 2, log $K_{OC}$ of bases (full circles) was plotted versus $pK_a$. Considering only bases with log $K_{OW} > 1$ (values on the right side of Fig. 6), which means that hydrophilic cations, for which we expect diffusion-limited transport, are excluded, $r^2$ between log $K_{OC}$ and $pK_a$ of bases increases from 0.07 to 0.55. For the lipophilic cations, the sorption increases with increasing $pK_a$ (significant correlation, $r < 0.001$), which indicates that strong cations are more competitive at binding to negatively charged sites in soil than weak cations. The $pK_a$ explains more than half ($r^2$ is 0.55) of the observed variance in the log $K_{OC}$ of lipophilic cations.

To account for the diffusion-limited transport from the bulk soil to the organic matter, we introduce a term $f$, which is defined as the ratio of the concentration in octanol ($C_{OCT}$) to the total concentration $C_{TOT}$: $C_{TOT} = C_{OCT} + C_W$ in the water ($C_W$) at pH 7. Varying from 0 to 1, $f$ is a descriptor for the fraction of compound in the lipophilic phase and thus may predict transport limitation of hydrophilic molecules by the organic (lipophilic) phase in soil.

$$f = \frac{C_{OCT}}{C_{TOT}} = \frac{C_{OCT}}{C_{OCT} + C_W} = \frac{C_W \cdot K_{GW}}{C_W \cdot K_{GW} + C_W} = \frac{K_{GW}}{K_{GW} + 1}$$

(22)

Reformulating the ionic contribution to sorption as a function of $pK_a$ and $f$, Equation 12 can be expressed as

$$\log K_{OC} = \log(\phi_a \cdot 10^{0.37 \log P_a + b} + \phi_{an} \cdot 10^{y \cdot f^4})$$

(23)

Keeping the default value of 12 for the $pK_{GW}$ of completely dissociated bases (four compounds in the training set and two in the validation), the nonlinear regression gave the following parameters:

$$\log K_{OC} = \log(\phi_a \cdot 10^{0.37 \log P_a + 1.70} + \phi_{an} \cdot 10^{y \cdot f^4})$$

(24)

The best fit is obtained with a pH$_{eq}$ at 4.5. Equation 24 is plotted in Figure 7, with some similarity to Equation 20 (Fig. 4). The differences in predictions, compared to Equation 20, are up to one log unit. The new regression gives higher log $K_{OC}$ for lipophilic strong bases ($log P_a > 2$, $pK_a > 8$) and for polar moderate bases ($log P_a < 2$, $pK_a$ between 5 and 10). It predicts lower log $K_{OC}$ for polar strong bases ($log P_a < 2$, $pK_a > 8$) and for lipophilic moderate bases ($log P_a > 3$, $pK_a$ between 5 and 10).

The accuracy of Equation 24 for the training set is good, similar to the previous regression Equation 20 (Fig. 8a); the
new regression is more accurate for the validation set (Fig. 8b). In addition, Equation 24 better considers the variance of the data ($r^2 = 0.76$ and 0.55 for the training and validation sets, respectively). The average error is smaller, too (RMSD = 0.38 and 0.51 for the training and validation sets, respectively). The two outliers acridine and benzo[f]quinoline are still as evident as before.

The performance of the regression for amphoters (Fig. 5f) is similar replacing the cationic contribution with the nonlinear term of Equation 24.

**Applicability**

Table 3 summarizes the equations we suggest for the prediction of log $K_{	ext{OC}}$ of ionizable compounds. All equations are applicable within the chemical space represented in Figure 1, which covers the whole pH range (0–12 for acids and 2 to >12 for bases). In the dataset, for acids $-2.18 < \log P_n < 8.50$; for bases $-1.66 < \log P_n < 7.03$.

For acids ($pK_a < 12$), Equation 15 (the combined regression) is preferred. For bases ($pK_a > 2$) Equation 24 gives the best estimates of log $K_{	ext{OC}}$. The suggested regression for amphoters (Eqn. 25) is composed of Equation 14 (anionic fraction), the second term of Equation 24 (cationic fraction) and Equation 21 (neutral fraction). This equation does not consider zwitterions. Thus, it is valid for $pK_{a,\text{base}} < pK_{a,\text{acid}}$ by at least two units so that the zwitterionic fraction is negligible.

**Selection of soil pH**

The variability of soil pH is one of the major reasons for the variation of experimental $K_{	ext{OC}}$ values of ionizable compounds [4]. The equations suggested in Table 3 theoretically can be adapted to varying soil pH, because the fraction of neutral and ionic species changes with pH (Eqn. 8). The pH was calibrated in the regressions to best fit the experimental data. The pH$_{\text{opt}}$ obtained in that way differs from the pH that was measured in the bulk soil solution (pH$_{\text{bulk}}$) during the experimental $K_{	ext{OC}}$-determination. Although the average pH$_{\text{bulk}}$ was 6.4 for acids and 5.8 for bases (Tables S1 and S2), the pH$_{\text{opt}}$ is 5.8 and 4.5 using Equations 15 and 24 for acids and bases, respectively. Thus, pH$_{\text{opt}} < $ pH$_{\text{bulk}}$ by 0.6 units for acids and 1.3 units for bases. Similarly, Bintein and Devillers [7] considered speciation of bases at two pH units lower than the bulk pH for the CF$_a$ (Eqns. 2 and 3).

The accurate pH for the prediction of the sorption of bases to organic carbon is much lower than the bulk pH of the soil. An explanation may be that the sorption of bases depends rather upon the dissociation ratio at the surface of organic colloids than on that in the bulk soil solution. Moving from the bulk towards the surface of organic colloids, the pH decreases because the pH of the organic fraction is usually acidic due to dissolved organic acids, and because sorption occurs at the surface of charged organic colloids, where the electrical potential is negative and the pH is lowered due to the Nernst potential of H$^+$ ions [14]

$$\text{pH}_{\text{surf}} = \text{pH}_{\text{bulk}} + \frac{F \cdot E_{\text{surf}}}{RT \cdot \ln(10)}$$

where pH$_{\text{surf}}$ is the pH obtained at the surface of the charged colloid, $F$ is the Faraday constant (96,484.56 C mol$^{-1}$), $R$ is the universal gas constant (8.314 J mol$^{-1}$ K$^{-1}$), $T$ is the absolute temperature (K), and $E_{\text{surf}}$ is the surface potential of the colloid (V). With $E_{\text{surf}} = -60$ mV, a pH decrease of one unit follows, alone due to the surface potential. Anions are repulsed by the negative charge, and cations are attracted. This may be the reason why this pH decrease was more evident for cationic compounds. However, a decrease of bulk soil pH to low values does not necessarily mean an increase of the sorption of cations.

**Table 3. Regressions suggested for the prediction of log $K_{	ext{OC}}$ of weak electrolytes (acids, bases, and amphoters); pH$_{\text{opt}}$ is the pH to input for the calculation of the neutral ($\phi_0$) and ionic fraction ($\phi_{\text{ion}}$); $\log P_n$ is the log $K_{\text{OW}}$ of the neutral molecule; $f$ is $K_{\text{OW}}/(K_{\text{OW}} + 1)$.**

<table>
<thead>
<tr>
<th>Equation</th>
<th>pH$_{\text{opt}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acids (Eqn. 15)</td>
<td>5.8</td>
</tr>
<tr>
<td>Bases (Eqn. 24)</td>
<td>4.5</td>
</tr>
<tr>
<td>Amphoters (Eqn. 25)</td>
<td>5</td>
</tr>
</tbody>
</table>

*Combination of Equations 14 (anionic fraction), 24 (second term only, for the cationic fraction) and 21 (neutral fraction).
because, at low soil pH (between 3 and 5), the carboxylic groups of the humic and fulvic acids in the soil organic matter will protonate and, therefore, fewer negatively charged sorption sites will be available [15].

**Limitations**

The result of the fit is a semiempirical estimation of the apparent soil–water partition coefficient normalized to organic carbon (K$_{OC}$), which may be determined by several processes such as lipophilic absorption and electrical adsorption involving dissociated and undissociated species. The dataset collected to derive the regressions covered a broad range of chemicals; therefore, the applicability of the equations is not limited to certain chemical classes. However, some other limitations shall be mentioned.

**Clay content.** The experimental K$_{OC}$ values were derived by dividing the measured distribution coefficients, K$_d$, by the fraction of organic carbon, f$_{OC}$, of the samples (Eqn. 4). The agreement of the proposed equations with the experimental log $K_{OC}$ confirms that this approach is valid and helpful. This may be because other constituents of bulk soil, such as clay or minerals, contribute less to the overall sorption. A second reason may be that the amount of clay is correlated with the amount of organic carbon [5]. Therefore, the proposed equations may fail if soil or sediment samples show an unusual ratio of organic carbon to clay (or other material that can adsorb), or if the adsorption to another fraction than organic carbon is exceptionally high. Examples are acridine and benzof[qu]inoline, two cations with delocalized charge, which were outliers from the regressions. This probably was due to their special affinity to the negatively charged clay [11].

**Complexation.** The effects of complexation were not addressed specifically. Organic ions can bind with metals or other ligands of opposite charge. The formation of a neutral complex alters speciation and sorption equilibria. The higher sorption of some acids dissociated at environmental pH may be explained by higher lipophilicity of the complexes formed.

**Sorption to sediments.** The experimental sorption data stem from soil samples. Nevertheless, there is no obvious reason why the equations should not be applicable for the prediction of sorption to sediment. The K$_{OC}$ of marine sediments may be affected by the high ionic strength of seawater (I is 0.7 M) [16]. Ionic strength is not considered in the regressions, because all experimental K$_{OC}$ refer to freshwater. Laak et al. [16] showed that ionic strength has a negligible effect in the range of natural freshwater (I < 0.02). Typical values of I of natural waters range from 0.01 in freshwater to 0.7 in seawater [17]. At higher ionic strength (salty waters), two effects can be observed: Increased competition with other cations for negative sorbing sites and enhanced complexation. The latter can impact differently on sorption due to diverse lipophilicity of the complexed molecule. Both effects were observed for the base tributyltin ([TBT]; pK$_a$ = 6.25 [18]) by Burton et al [19].

The sorption of TBT increased with increasing ionic strength when the pH was above the pK$_a$ (i.e., when the neutral fraction dominated), due to the formation of TBT-Cl, which is more lipophilic than TBT-OH (TBT-Cl: log $P_{n}$ = 4.76; TBT-OH: log $P_{n}$ = 4.2 [18]). When the cationic form dominated, i.e., when pH was below pK$_a$, the sorption of TBT decreased with increasing I, due to competition with other cations for the sorption sites.

**Polar interactions** influence the sorption of neutral polar molecules, but the K$_{OW}$ is not a good predictor for polar interactions [20,21]. Linear solvation energy relationships based on molecular and environmental media properties [21] and models based on molecular fragments [22] have improved the predictive power and enlarged the applicability of soil–water partitioning quantitative structure-activity relationships to hydrophilic polar compounds. However, they do not consider ionic interactions. Although not applied in the present study, these methods might be used to predict the sorption of the neutral molecule (Eqn. 21) of polar weak electrolytes. This could refine the prediction of K$_{OC}$.

**CONCLUSION**

Using log $P_{n}$ and pK$_a$ as predictors, the sorption of organic electrolytes with cationic dissociating groups (cations and amphoters) to organic carbon could be predicted successfully. Conversely, it was difficult to develop predictive equations for acids with pK$_a$ < 6 based on the same predictors. The sorption of anions generally was low (K$_{OC}$ between 10 and 300 L/kg) with a few exceptions, which cannot be explained from pK$_a$ and log $P_{n}$.

The equations suggested in Table 3 are applicable over the whole pK$_a$ range and are useful in particular for relatively strong bases and amphoters, for which no predictive method specifically had been developed so far. As was shown, prediction methods that consider the sorption of electrolytes by treating neutral and ionic fraction separately can provide good estimates of the log K$_{OC}$ and probably are superior to methods that merge both fractions without considering the differences between neutral compounds and ions.

**SUPPORTING INFORMATION**

Table S1. Dataset of acids: Dissociation constant (pK$_a$) and log $P_{n}$ (ACD); experimental log K$_{OC}$ and average bulk soil pH of the experiments (see reference); calculated log K$_{OC}$. (Eqn. 15 with pH$_{opt}$ = 5.8).

Table S2. Dataset of bases: Dissociation constant (pK$_a$) and log $P_{n}$ (ACD, unless specified); experimental log K$_{OC}$ and average bulk soil pH of the experiments (see reference); calculated log K$_{OC}$. (Eqn. 20, pH$_{opt}$ is 6 and Eqn. 24, pH$_{opt}$ is 4.5).

Table S3. Dataset of amphoters: Dissociation constants (pK$_{wa}$ basic and pK$_{wa}$ acid) and log $P_{n}$ (ACD); experimental log K$_{OC}$ (see reference); calculated log K$_{OC}$. (Eqn. 25, pH$_{opt}$ is 5.0).

All found at DOI: 10.1897/07-583.S1 (52 KB PDF).

**Acknowledgement**—The present study received financial support from the European Union Sixth Framework Program of Research, Thematic Priority 6 (Global change and ecosystems), contract GOCE-CT-2007-036976, project 2-FUN, and contract GOCE-037017, project Optimized Strategies for Risk Assessment of Chemicals. Support for the present study also was provided through a PhD grant of the Technical University of Denmark for A. Franco.

**REFERENCES**


