Estimating the fates of organic contaminants in an aquifer using QSAR
Seung Joo Lim and Peter Fox

ABSTRACT

The quantitative structure activity relationship (QSAR) model, BIOWIN, was modified to more accurately estimate the fates of organic contaminants in an aquifer. The predictions from BIOWIN were modified to include oxidation and sorption effects. The predictive model therefore included the effects of sorption, biodegradation, and oxidation. A total of 35 organic compounds were used to validate the predictive model. The majority of the ratios of predicted half-life to measured half-life were within a factor of 2 and no ratio values were greater than a factor of 5. In addition, the accuracy of estimating the persistence of organic compounds in the sub-surface was superior when modified by the relative fraction adsorbed to the solid phase, $1/R_f$, to that when modified by the remaining fraction of a given compound adsorbed to a solid, $1/C_0 f_s$.

INTRODUCTION

Since the first report of incomplete removal of steroid hormones during wastewater treatment in the 1960s, many investigators have shown that pharmaceuticals and personal care products (PPCPs) were present in the aquatic environment (Stumm-Zollinger & Fair 1965; Ternes 1998; Kolpin et al. 2002; Westerhoff et al. 2005; Verlicchi et al. 2012). Of greatest concern is that the effects of PPCPs released into the environment to humans and/or ecosystems are still dubious. Specifically, in spite of even low level of PPCPs, interactions between enzymes and pharmaceutical compounds may be important (Monteiro & Boxall 2010). Thus, several investigators showed the results for the treatment of micropollutants using advanced oxidizing processes (Baumgarten et al. 2007; Lee et al. 2012). Several quantitative structure activity relationship (QSAR) tools for estimating persistence ($P$), bioaccumulation ($B$), and toxicity ($T$) of organic contaminants have been developed including EPI Suite (US EPA 2005). EPI Suite primarily consists of two QSAR models, which are BIOWIN and AOPWIN, and it has been developed and improved since the 1980s. Nonetheless, many investigators reported several flaws and limitations of BIOWIN (Kühne et al. 2006; Niederer & Goss 2008). Although the accuracy for estimating the persistence of organic compounds using EPI Suite was improved, it is necessary that QSAR models be developed to predict the fates of contaminants in the sub-surface. This is particularly true as reclaimed water is increasingly being used to augment groundwater supplies. In this study, the predictions from BIOWIN were modified using oxidation and sorption to more accurately estimate the fates of contaminants in an aquifer.

METHODOLOGY

Modification of predictions from EPI Suite to estimate the persistence of organic contaminants in an aquifer

Even though EPI Suite has been widely used for estimating the persistence ($P$), bioaccumulation ($B$), and toxicity ($T$) of organic compounds for some years, it does not consider important environmental factors such as sorption or oxidation of compounds by co-metabolic reactions. It is necessary that EPI Suite, based on BIOWIN and AOPWIN, should be modified to obtain more accurate predictions of contaminants moving in an aquifer. Sorption plays an important role in determining the movement and biodegradation of organic contaminants in the sub-surface. Hydrophobic compounds are slowly released due to
sorption/desorption in soils. Subsequently, the biodegradability of released organic contaminants is highly affected by the activity of microorganisms. This implies that the persistence of organic contaminants should be modified by microbial oxidation and sorption in the sub-surface. A compound’s susceptibility to microbial oxidation by cometabolism is not considered in BIOWIN, but these reactions are important during sub-surface transport. Considering diffusion limitations in the aqueous phase are more pronounced as compared to the gas phase, a weighting factor between air and aqueous phase reaction rates can be correlated. In order to generate the weighting factor, the compounds used were taken from databases of current commonly used pharmaceutical compounds including the Top Branded Rx Drugs by Units Dispensed (2007), Top Generic Rx Drugs by Units Dispensed (2007), and Top Hospital Rx Products by Wholesale Acquisition Costs (2007). The weighting factor between air and aqueous phase reactions was 1,940.3 (Lim & Fox 2012).

\[ \log K_{OX \text{ [water]}} = \log K_{OX \text{ [air]}} - 3.29 \quad (1) \]

where, \( K_{OX} \): the first-order oxidation reaction rate constant [1/d].

The half-life prediction from BIOWIN was modified to include the potential for oxidation during sub-surface transport. This modification used the results from AOPWIN with the weighting factor to adjust for aqueous phase reaction rates. The following equations show that the predicted half-life from BIOWIN was modified to include both biodegradation and oxidation.

\[ C = C_o e^{-[k_1 + k_3]t} \quad (2) \]

\[ \ln(0.5) = -[k_1 + k_3]t \quad (3) \]

\[ t_o = \frac{-\ln(0.5)}{(k_1 + k_3)} \quad (4) \]

where, \( k_1 \): biodegradation reaction rate constant in water phase [1/T], \( k_3 \): oxidation reaction rate constant in water phase [1/T], \( k_2 \): oxidation reaction rate constant in air phase/1,940.5 [1/T], 1,940.3 = weighting factor, \( t_o \): predicted half-life including oxidation [T].

The predictions from EPI Suite were modified to include the effects of sorption using the relationships suggested by Heidler & Halden (2008) and Bouwer (1991). According to the research of Heidler & Halden (2008), the fraction of a given compound adsorbed to solids, \( f_s \), is a function of \( K_{oc} \) and a fitting parameter \( (4.21 \times 10^{-5}) \) (Equation (5)).

\[ f_s = \frac{K_{oc} \times 4.21 \times 10^{-5}}{(1 + K_{oc} \times 4.21 \times 10^{-5})} \quad (5) \]

Thus, \( 1 - f_s \) is the fraction of a given compound remaining in solution (Equation (6)).

\[ 1 - f_s = \frac{1}{(1 + K_{oc} \times 4.21 \times 10^{-5})} \times \frac{f_{oc}}{2.87 \times f_{oc} \times (R_f - 1) \times 4.21 \times 10^{-5}} \quad (6) \]

Bouwer (1991) showed the retardation factor, \( R_f \), as described in Equation (7). Most aquifers used for groundwater recharge with reclaimed water are composed of sands and/or gravels. For a sand aquifer system, the retardation factor \( R_f \) for a given compound, can be defined as follows:

\[ R_f = \frac{v_{gw}}{v_{sp}} = 1 + \frac{\rho b_{k_d}}{\theta} = 1 + 2.87K_d = 1 + \frac{\rho f_{oc}K_{oc}}{\theta} = 1 + 0.02K_{oc} \quad (7) \]

where, \( R_f \): retardation factor (dimensionless); \( v_{gw} \): the velocity of groundwater (cm/d); \( v_{sp} \): the velocity of contaminant species (cm/d); \( \rho_b \): bulk density (1.62 kg/L (sand aquifer), Ohio EPA 2005); \( \theta \): water content (0.5654 water volume L/total volume L (sand aquifer), Ohio EPA 2005); \( K_d \): distribution coefficient (L/kg); \( K_{oc} \): sorption coefficient (L/kg); \( f_{oc} \): fraction of OC (0.0071 L/kg C (sand aquifer), Ohio EPA 2005).

As shown in Equations (5) and (7), the fraction of compounds adsorbed to a solid, \( f_s \), is proportional to the retardation factor, \( R_f \). Equation (7) can be rearranged as follows:

\[ R_f = 1 + \frac{\rho b_{k_d}^2}{\theta} \left( \frac{1 - [1 - f_s]}{4.21 \times 10^{-5} \times [1 - f_s]} \right) \]

\[ = 1 + 1.4 \times 10^{-4} \left( \frac{1 - [1 - f_s]}{4.21 \times 10^{-5} \times [1 - f_s]} \right) \quad (8) \]

\[ \frac{1}{R_f} \left( 1 - V_{sp} \right) = 1 - \frac{V_{gw}}{R_f} = 1 - \left( 1 - \frac{1}{R_f} \right) \]

\[ = 1 - \frac{V_{gw}}{V_{gw}} \quad (9) \]
In Equation (7), the fraction of a compound adsorbed to a solid is a function of $K_{oc}$ and soil properties (bulk density, fraction of organic carbon, and water content). The effects of sorption on fate of a compound in an aquifer can be predicted based on the fraction of compound adsorbed to a solid. In other words, this implies that the relative fraction adsorbed to the solid phase, $1/R_f$, is proportional to the quantity of compound available for biodegradation, $1/C_{fs}$, to predict the half-life from BIOWIN. The predicted half-life from BIOWIN was modified to include sorption in two different manners. Equation (10) modifies the prediction based on the reduction in aqueous phase concentration while Equation (11) modifies the prediction based on the increased time required for transport in an aquifer.

predicted half-life to include sorption and oxidation ($t_s$) =
$$(1 - f_s) \times \left( -\frac{\ln(0.5)}{k_1} \right)$$

predicted half-life to include sorption and oxidation ($t_s$) =
$$\frac{1}{R_f} \times \left( -\frac{\ln(0.5)}{k_1} \right)$$

Consequently, the overall expression to include both the effects of sorption and oxidation may be expressed as follows:

predicted half-life to include sorption and oxidation ($t_{so}$) =
$$(1 - f_s) \times \left( -\frac{\ln(0.5)}{(k_1 + k_3)} \right)$$

\[ t_{so} = \frac{1}{R_f} \times \left( -\frac{\ln(0.5)}{(k_2 + k_3)} \right) \]  

RESULTS AND DISCUSSION

Correlation between $1/R_f$ and $1 - f_s$ to predict half-lives of contaminants in the sub-surface

As shown in Equations (10)–(13), the predicted half-life to include sorption or to include both the effects of sorption and oxidation has two equations due to the different sorption expressions. The correlation between $1/R_f$ and $1 - f_s$ is shown in Figure 1. Although both variables imply the effect of sorption, in Equations (12) and (13), the values of $1/R_f$ were relatively underestimated.

As shown in Equations (6) and (7), both $1 - f_s$ and $1/R_f$ were a function of $K_{oc}$. Equations (6) and (7), $R_f(1 - f_s)$ can be rearranged as follows:

$$R_f(1 - f_s) = \frac{1 + 0.02K_{oc}}{(1 + 4.21 \times 10^{-9} \times K_{oc})}$$

Equation (14) shows sorption is highly dependent upon $K_{oc}$ and the difference between $1 - f_s$ and $1/R_f$ was determined by the $K_{oc}$ of each contaminant. In order to compare the effect of $K_{oc}$ on the predictions from BIOWIN including sorption, the logarithmic ratio of $1 - f_s$ to $1/R_f$ was plotted as a function of log $K_{oc}$ (Figure 2).

![Figure 1](https://iwaponline.com/wst/article-pdf/68/1/153/439990/153.pdf)

Relationship between the two expressions.
When log $K_{oc}$ approached 6.2, log $R_f(1 - f_s)$ converged to 0.53 as shown in Figure 2, and 0.53 was the maximum value. The values of log $R_f(1 - f_s)$ sharply increased as the log $K_{oc}$ value increased from 3.7 to 6.2. This evidently shows that $1/R_f$ was underestimated as compared to $1/C_0f_s$. The value of $1/C_0f_s$ was 3.43 times as high as that of $1/R_f$ for a log $K_{oc}$ greater than 6.2. In addition, the difference between $1/C_0f_s$ and $1/R_f$ was negligible when log $K_{oc}$ of a contaminant was greater than 6.2 because log $R_f(1 - f_s)$ converged to 0.53. In other words, this implies that the difference between Equation (12) and Equation (13) was significant for compounds with log $K_{oc}$ ranging from 3.7 to 6.2. On the other hand, when log $K_{oc}$ of a contaminant was less than 3.7, it was critical to understand that $1/R_f$ was underestimated. Given that log $K_{oc}$ increased in Equation (14), log $R_f(1 - f_s)$ is a sigmoidal curve. Thus, the value of log$(1 - f_s)$ is more than affected by that of log$(1/R_f)$, because log$(1/R_f)$ was underestimated (Figure 1). On the other hand, when log $K_{oc}$ decreased, log$(1/R_f) + log(1 - f_s)$ approached zero. This implies that the relationship between the two (sorptions) was the same in Equations (12) and (13).

### Comparison with modified half-life equation

The half-life predictions, $t_{so}$, when using $1 - f_s$ in Equation (12) were compared to those when using $1/R_f$ in Equation (13). Comparisons of chemicals were performed on the results for half-lives reported in the literature for subsurface transport. The summary of the comparison results is shown in Table 1 where the reference half-lives of 35 compounds were based on experimental results. The compounds were chosen to represent emerging micropollutants such as pharmaceutical drugs, pesticides/herbicides, pesticides/herbicides,

---

**Table 1** | Comparison of a modified half-life equation using some organic compounds

<table>
<thead>
<tr>
<th>Substance</th>
<th>log $K_{oc}$</th>
<th>BIOWIN half-life (d)$^b$</th>
<th>Reference half-life (d)$^c$</th>
<th>$t_{so}$ (d)$^d$</th>
<th>$t_{so}$/reference half-life$^e$</th>
<th>$t_{so}$/reference half-life$^g$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetaminophen</td>
<td>1.79</td>
<td>15</td>
<td>3.1</td>
<td>14.8</td>
<td>14.7</td>
<td>4.76</td>
<td>4.75</td>
</tr>
<tr>
<td>Atrazine</td>
<td>2.35</td>
<td>60</td>
<td>60</td>
<td>55.1</td>
<td>53.9</td>
<td>0.91</td>
<td>0.90</td>
</tr>
<tr>
<td>α-BHC</td>
<td>3.45</td>
<td>180</td>
<td>32</td>
<td>160.8</td>
<td>127.9</td>
<td>5.02</td>
<td>4.00</td>
</tr>
<tr>
<td>β-BHC</td>
<td>3.45</td>
<td>180</td>
<td>34.5</td>
<td>160.8</td>
<td>127.9</td>
<td>3.69</td>
<td>2.94</td>
</tr>
<tr>
<td>γ-BHC</td>
<td>3.45</td>
<td>180</td>
<td>32</td>
<td>160.8</td>
<td>127.9</td>
<td>5.02</td>
<td>4.00</td>
</tr>
<tr>
<td>δ-BHC</td>
<td>3.45</td>
<td>180</td>
<td>57</td>
<td>160.8</td>
<td>127.9</td>
<td>2.82</td>
<td>2.24</td>
</tr>
<tr>
<td>Bisphenol A</td>
<td>4.58</td>
<td>37.5</td>
<td>30</td>
<td>12.7</td>
<td>5.1</td>
<td>0.42</td>
<td>0.17</td>
</tr>
<tr>
<td>Carbamazepine</td>
<td>3.12</td>
<td>37.5</td>
<td>82</td>
<td>22.6</td>
<td>20.0</td>
<td>0.27</td>
<td>0.24</td>
</tr>
</tbody>
</table>

(Continued)
polycyclic aromatic hydrocarbons (PAHs), and halogenated hydrocarbons. These compounds contain a variety of different structures and functions and have been widely used and/or investigated. The majority of the ratios of $t_{so}$ to reference half-lives were within a factor of 2 and no values were more than a factor of 5 different. The average ratio of $t_{so}$ to reference half-life was $1.88 \pm 1.57$ using Equation (12), whereas the average ratio was $1.42 \pm 1.50$ using Equation (13). This implies that BIOWIN can be modified using sorption and oxidation to improve half-life predictions for sub-surface transport. Nendza (1998) stated that organic compounds with a log $K_{oc}$ greater than 3.5 can be categorized as hydrophobic. Considering the values of log $R_f(1-f_s)$ sharply increased for log $K_{oc}$ values between 3.7 and 6.2, the effects of hydrophobicity on the modification of prediction from BIOWIN were also significant. For a 95% confidence interval, the ratio of $t_{so}$ to reference half-life using Equation (12) was 1.36–2.40. Meanwhile, the confidence interval using Equation (13) was 1.00–1.85. This demonstrates that

<table>
<thead>
<tr>
<th>Substance</th>
<th>log $K_{oc}$</th>
<th>BIOWIN half-life (d)$^a$</th>
<th>Reference half-life (d)$^b$</th>
<th>$t_{so}$ (d)$^c$</th>
<th>$t_{so}$/reference half-life$^d$</th>
<th>$t_{so}$/reference half-life$^e$</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDD</td>
<td>5.08</td>
<td>180</td>
<td>75</td>
<td>28.7</td>
<td>9.5</td>
<td>0.38</td>
<td>0.13</td>
</tr>
<tr>
<td>DDE</td>
<td>5.08</td>
<td>180</td>
<td>75</td>
<td>26.9</td>
<td>8.9</td>
<td>0.35</td>
<td>0.12</td>
</tr>
<tr>
<td>DDT</td>
<td>5.24</td>
<td>180</td>
<td>50</td>
<td>21.2</td>
<td>6.8</td>
<td>0.42</td>
<td>0.14</td>
</tr>
<tr>
<td>Diclofenac</td>
<td>2.66</td>
<td>37.5</td>
<td>7.7</td>
<td>28.4</td>
<td>27.1</td>
<td>3.68</td>
<td>3.52</td>
</tr>
<tr>
<td>Dilantin</td>
<td>3.17</td>
<td>37.5</td>
<td>75</td>
<td>34.6</td>
<td>30.3</td>
<td>0.46</td>
<td>0.40</td>
</tr>
<tr>
<td>Ethynylestradiol</td>
<td>4.65</td>
<td>60</td>
<td>5.2</td>
<td>15.3</td>
<td>5.9</td>
<td>2.94</td>
<td>1.14</td>
</tr>
<tr>
<td>Endrin</td>
<td>4.30</td>
<td>180</td>
<td>31</td>
<td>90.3</td>
<td>42.7</td>
<td>2.91</td>
<td>1.38</td>
</tr>
<tr>
<td>Erythromycin</td>
<td>2.75</td>
<td>180</td>
<td>20</td>
<td>39.1</td>
<td>37.0</td>
<td>1.96</td>
<td>1.85</td>
</tr>
<tr>
<td>17β-Estradiol</td>
<td>4.19</td>
<td>37.5</td>
<td>3.7</td>
<td>18.6</td>
<td>9.5</td>
<td>5.04</td>
<td>2.58</td>
</tr>
<tr>
<td>Estriol</td>
<td>3.17</td>
<td>37.5</td>
<td>17.3</td>
<td>34.6</td>
<td>30.3</td>
<td>2.00</td>
<td>1.75</td>
</tr>
<tr>
<td>Estrone</td>
<td>4.38</td>
<td>37.5</td>
<td>42.2</td>
<td>15.3</td>
<td>6.9</td>
<td>0.36</td>
<td>0.16</td>
</tr>
<tr>
<td>Fluoxetine</td>
<td>4.97</td>
<td>60</td>
<td>6.6</td>
<td>10.9</td>
<td>3.7</td>
<td>1.66</td>
<td>0.57</td>
</tr>
<tr>
<td>Gemfibrozil</td>
<td>2.64</td>
<td>37.5</td>
<td>75</td>
<td>32.1</td>
<td>30.7</td>
<td>0.43</td>
<td>0.41</td>
</tr>
<tr>
<td>Heptachlor</td>
<td>4.62</td>
<td>180</td>
<td>82.5</td>
<td>21.6</td>
<td>8.5</td>
<td>0.26</td>
<td>0.10</td>
</tr>
<tr>
<td>Ibuprofen</td>
<td>2.63</td>
<td>15</td>
<td>7.2</td>
<td>14.6</td>
<td>14.0</td>
<td>2.05</td>
<td>1.95</td>
</tr>
<tr>
<td>Iopromide</td>
<td>1.00</td>
<td>60</td>
<td>17.3</td>
<td>50.5</td>
<td>50.4</td>
<td>2.91</td>
<td>2.91</td>
</tr>
<tr>
<td>Ketoprofen</td>
<td>2.59</td>
<td>15</td>
<td>8.0</td>
<td>14.7</td>
<td>14.1</td>
<td>1.84</td>
<td>1.77</td>
</tr>
<tr>
<td>Metolachlor</td>
<td>4.43</td>
<td>180</td>
<td>90</td>
<td>57.6</td>
<td>25.1</td>
<td>0.64</td>
<td>0.28</td>
</tr>
<tr>
<td>Methoxychlor</td>
<td>2.69</td>
<td>60</td>
<td>26</td>
<td>50.6</td>
<td>48.2</td>
<td>1.95</td>
<td>1.86</td>
</tr>
<tr>
<td>Musk ketone</td>
<td>3.40</td>
<td>60</td>
<td>180</td>
<td>54.1</td>
<td>43.9</td>
<td>0.30</td>
<td>0.24</td>
</tr>
<tr>
<td>Naproxen</td>
<td>2.53</td>
<td>15</td>
<td>7.6</td>
<td>15.7</td>
<td>13.2</td>
<td>1.80</td>
<td>1.74</td>
</tr>
<tr>
<td>Oleandomycin</td>
<td>1.36</td>
<td>60</td>
<td>27</td>
<td>33.4</td>
<td>33.3</td>
<td>1.24</td>
<td>1.23</td>
</tr>
<tr>
<td>Phenanthrene</td>
<td>4.22</td>
<td>60</td>
<td>8.7</td>
<td>33.9</td>
<td>16.9</td>
<td>3.90</td>
<td>1.95</td>
</tr>
<tr>
<td>Pyrene</td>
<td>4.74</td>
<td>60</td>
<td>33</td>
<td>15.9</td>
<td>5.9</td>
<td>0.48</td>
<td>0.18</td>
</tr>
<tr>
<td>Sulfamethoxazole</td>
<td>2.41</td>
<td>37.5</td>
<td>19</td>
<td>27.2</td>
<td>26.5</td>
<td>1.43</td>
<td>1.40</td>
</tr>
<tr>
<td>Triclosan</td>
<td>4.37</td>
<td>60</td>
<td>28.9</td>
<td>28.9</td>
<td>13.1</td>
<td>1.00</td>
<td>0.45</td>
</tr>
<tr>
<td>Trimethoprim</td>
<td>2.86</td>
<td>60</td>
<td>80</td>
<td>36.6</td>
<td>34.2</td>
<td>0.46</td>
<td>0.43</td>
</tr>
</tbody>
</table>

$^a$Estimated by EPI Suite (ver. 4.1).

$^b$Half-life prediction from BIOWIN.

$^c$Reference half-life from reference.

$^d$Modified half-life using Equation (12).

$^e$Modified half-life using Equation (13).

$^f$Ratio of modified half-life using Equation (12) to reference half-life.

$^g$Ratio of modified half-life using Equation (13) to reference half-life.
more accurate estimates of persistence of organic contaminants in the sub-surface may be obtained using Equation (13).

CONCLUSIONS

The predictions from BIOWIN were successfully modified by sorption and oxidation. The modified half-life equation included the effects of sorption, biodegradation, and oxidation. The majority of the ratios of the modified half-life to reference half-life were within a factor of 2 of the experimental values. The modified predictions from BIOWIN were more accurate when the relative fraction adsorbed to solid phase, $1/R_b$, was used to describe the effects of sorption.

ACKNOWLEDGEMENTS

This research was supported by the WaterReuse Foundation project entitled ‘Tools to Assess and Understand the Relative Risks of in Indirect Potable Reuse and Aquifer Storage Recovery Projects’. The authors are indebted to the co-pro-ject investigators on the project including Louisa Flander, Andrew Speirs-Bridge, Grant Skrepnak, Jeffrey Soller, and Margaret Nellor.

REFERENCES


US EPA 2005 Pollution prevention (P2) framework. EPA-748-B-04-001.


First received 2 December 2012; accepted in revised form 27 February 2013