Chemical Toxicity Correlations for Several Fish Species Based on the Abraham Solvation Parameter Model

Kaci R. Hoover,† William E. Acree, Jr.,*,† and Michael H. Abraham‡

Department of Chemistry, University of North Texas, P.O. Box 305070, Denton, Texas 76203-5070, and Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, United Kingdom

Received June 17, 2005

The Abraham solvation parameter model is used to construct mathematical correlations for describing the nonspecific aquatic toxicity of organic compounds to the fathead minnow, guppy, bluegill, goldfish, golden orfe, and high-eyes medaka. The derived mathematical correlations describe the observed published toxicity data to within an overall average standard deviation of approximately 0.28 log units. In the case of ester solutes, the descriptions were improved by introducing an indicator variable into the basic model. Derived correlations can be used to estimate aquatic toxicities of organic chemicals to the six fish species studied and to help in identifying compounds whose toxic mode of action might involve chemical specific reactivity, rather than nonpolar or polar narcosis. A principal component analysis of the correlation equations shows that the water—octanol system is a poor model for nonspecific aquatic toxicity but that the water—isobutanol and water—pentanol systems are much better models.

Introduction

Every year, new chemicals are produced and/or identified as the result of new industrial or natural biological processes. Not all of these compounds are safe. Many exhibit adverse environmental effects. Aquatic toxicity is one of a batch of texts designed to assess the damage that would result if a particular chemical were to be released or were to make its way into our natural waterways. Standard test methods and experimental protocols have been established for determining the median mortality lethal concentration, LC₅₀, for evaluating the chronic toxicity, for determining decreased population growth, and for quantifying developmental toxicity at various life stages for several different aquatic organisms. Experimental determinations are often very expensive and time-consuming as several factors may need to be carefully controlled in order to adhere to the established, recommended experimental protocol. For example, published studies (1–6) have shown that the aquatic toxicity of select polycyclic aromatic hydrocarbons (PAHs) can be enhanced by ultraviolet radiation.

Aquatic toxicity data are available for relatively few organic, organometallic, and inorganic compounds. To address this concern, researchers have developed predictive methods as a means to estimate toxicities in the absence of experimental data. Expressions have been developed for predicting aquatic toxicity from water to octanol partition coefficients (7–11), from theoretical indices/descriptors calculated from structural information (11–15), from indices/descriptors of experimental origin (16–19), or from group contribution concepts (20, 21). Derived correlations have shown varying degrees of success in their ability to predict the aquatic toxicity of different chemical compounds. In general, predictive methods are much better at estimating the aquatic toxicities of compounds that act through noncovalent or nonspecific modes of action. Nonpolar narcosis and polar narcosis are two such modes of nonspecific action. Most industrial organic compounds have either a nonpolar or polar narcotic mode of action, which lacks covalent interactions between toxicant and organism (22). Predictive methods are generally less successful in predicting the toxicity of compounds whose action mechanism involves electrophilic/nucleophilic covalent reactivity or receptor-mediated functional toxicity.

Subclasses within the broad narcosis classification include nonpolar, polar, ester, and amine narcotics. Nonpolar narcotic toxicity is often referred to as “baseline” or minimum toxicity. Polar narcotics exhibit effects similar to nonpolar narcotics; however, their observed toxicities are slightly more than “baseline” toxicity. Esters and alkylamines both exhibit narcotic toxicities greater than baseline toxicity. Aromatic amines have been classified in aquatic toxicology as polar narcotics (22–25). Published studies (18, 26) have shown that for certain aquatic organisms the toxicity of select amines/esters is similar to that of polar narcotics. Diethyl phthalate and ethyl 4-aminobenzoate are two such examples in that their toxicity toward the fathead minnow and guppy is more in line with that of a polar narcotic molecule, rather than that of an ester and/or alkylamine narcotic. Other esters/amines exhibit much higher toxicities than expected of polar narcotic compounds. One finds in reading the published literature (26–28) that the mode of toxic action of a given ester/amine may be listed as polar narcosis or ester/amine narcosis, depending upon the aquatic organism and predictive method under consideration.

In the present study, we reexamine the applicability of the Abraham solvation parameter model in regards
to its ability to correlate the observed toxicity data of nonpolar and polar compounds on fish. Specific species of fish studied are fathead minnow (Pimephales promelas), guppy (Poecilia reticulata), bluegill (Lepomis macrochirus), golden orfe (Leuciscus idus melanotus), goldfish (Carassius auratus), and medaka high-eyes (Oryzias latipes). For the latter species, two end points of lethal concentration are considered, both the 48 and the 96 h 50% mortality log LC₅₀ values. Kamlet et al. (16) previously showed that the toxicities of 32 organic nonelectrolytes to the golden orfe fish were well-correlated by

$$\log \text{LC}_{50} (\text{mM}) = -3.19 + 3.29(V/100) + 1.14\pi^* - 4.60\beta + 1.52\alpha_m$$  (1)

an earlier version of our current solvation parameter model, with the number of compounds being 32, the correlation coefficient squared being 0.966, and the standard deviation being 0.19. In eq 1, the independent solute/toxicant property is the logarithm of the 48 h 50% mortality lethal millimolar concentration. The dependent variables are the solute’s molar volume, V, and the solvatochromic parameters π*, β, and α_m are measures of the solute/toxicant dipolarity/polarizability, hydrogen bond acceptor basicity, and hydrogen bond donor acidity, respectively. Equation 1 underpredicted the toxicities of the esters by 0.3–1.1 log units. The authors noted that the enhanced toxicities of esters were consistent with a mechanism involving in vivo hydrolysis. In some aquatic organisms, esters may hydrolyze rapidly to give a greatly enhanced toxicity, while in other organisms in vivo hydrolysis may precede at a much slower rate so that the compound’s toxicity results from nonpolar/polar narcosis, rather from a reactive mechanism.

Abraham and Rafols (17) subsequently correlated the Overton data (29, 30) and data from the publication of Lipnick (31) for tadpole (Rana temporaria) narcosis with the revised, modern version of the solvation parameter model

$$\log C_{narc} = 0.609 + 0.866E - 0.347S - 0.174A - 2.808B^* + 3.054V$$  (2)

where C_{narc} is the narcotic concentration of the solute (mol dm⁻³), E is the excess solute molar refraction, S is the solute dipolarity/polarizability, A and B* denote the solute overall or effective hydrogen bond acidity and basicity, respectively, and V is the solute McGowan volume. (For notational simplicity, we have used the newest descriptor abbreviations.) Equation 2 had N = 84, R² = 0.951, SD = 0.246, and F = 351.0. Here and elsewhere, N is the number of data points, that is the number of solutes, R denotes the correlation coefficient, SD is the standard deviation, and F corresponds to the Fischer F-statistic. Larger data sets of more diverse solutes were obtained by compiling experimental toxicity data of several tadpole species (R. temporaria, Rana pipiens, and Xenopus laevis). The resulting equation

$$\log C_{narc} = 0.582 + 0.770E - 0.696S - 0.243A - 2.592B^* + 3.343V$$  (3)

had N = 114, R² = 0.906, SD = 0.337, and F = 217. The authors assigned a likely experimental error/uncertainty of 0.20 log units to the experimental −log C_{narc} values based on replicate published toxicity data for ethanol, 1-propanol, and 1-butanol determined by independent laboratories. The derived correlations described the observed toxicity data on tadpoles to approximately within the estimated experimental error. More recently, Gunatilleka and Poole (18, 19) applied the Abraham solvation parameter equation to the aquatic toxicities of several fish, zooplankton, algae, and bacteria. Toxicity end points were taken as the acute lethal concentration for median mortality (i.e., fish), effective concentration for 50% immobilization (i.e., water flea), 50% inhibition of cell growth (i.e., Tetrahymena pyriformis), or diminution of bioluminescence (i.e., Vibrio fischeri). Large deviations between the observed toxicity and the calculated values based on the correlations of Gunatilleka and Poole suggested that the compound’s mode of action might be different than either nonpolar or polar narcosis. Yu and co-workers (32) published a comparative study examining the applicability of four quantitative structure–activity relationship (QSAR) and linear free energy relationship (LFER) models in correlating the toxicity of aromatic compounds to Daphnia magna. One of the four models considered was the earlier version of the Abraham solvation parameter model. Only 43 compounds were considered in this latter study.

Our study differs from the prior work of Gunatilleka and Poole in that six species of fish are studied as opposed to three, much larger databases of 196 (ours) vs 119 (theirs) compounds and 148 (ours) vs 110 (theirs) compounds are used for the fathead minnow and guppy, respectively, the three largest databases (fathead minnow, guppy, and bluegill) are divided into separate training and test sets to validate the robustness of the derived correlations, and ester narcotics are examined in much greater detail. For esters that exhibit “excess toxicity”, we also determined whether the correlations could be improved by introducing an indicator variable into the descriptive equation. Except for diethyl phthalate and ethyl 4-aminobenzoate, esters were excluded from the correlations developed by Gunatilleka and Poole. In addition, we have analyzed the derived correlations using principal component analysis to determine which particular water–organic solvent system(s) best mimics chemical toxicity to fish.

**Experimental Procedures**

Toxicity data for the six species of fish were compiled from several published papers (see the Supporting Information) and from values contained in the ACQUIRE (33) and PAN (34) (Pesticide Area Network) Databases. ACQUIRE is part of the ECOTOX database, which was created and is maintained by the U.S. Environmental Protection Agency, the Office of Research and Development, and the National Health and Environmental Effects Research Laboratory’s Mid-Continental Ecology Division. Not all of the experimental data found were selected for inclusion in our databases. We rejected experimental data that were determined using compounds of less than desired chemical purity, data that were marked as “an outlier” in either the ACQUIRE or the PAN Databases, and data for compounds whose toxic mechanism was believed to be one of chemical reactivity, electro(nucleo)philic covalent reactivity, or receptor-mediated functional toxicity based on prior toxicity studies involving the compound. In instances where multiple entries were found in the ACQUIRE and PAN Databases, we either took the simple arithmetic average of the values found or, if the fish specimens studied were of great size disparity, we
selected the value for the fish specimen whose size closely matched the other values in our data set(s). To the extent possible, we tried to keep the fish size as constant as possible within each given data set. Our search of the chemical literature yielded log \( \text{LC}_{50} \) data for a total of 198 compounds (plus values for an additional 17 esters) for fathead minnow, a total of 148 compounds for guppy, a total of 69 compounds for bluegill, a total of 52 compounds (plus values for an additional 10 esters) for golden orfe, a total of 44 compounds (96 h end point) and 49 compounds (48 h end point) for medaka high-eyes, and a total of 51 compounds for goldfish. The experimental values are compiled in Tables 1–7 of the Supporting Information. Toxicity data were found for other species of fish [i.e., 18 compounds for zebra danio (\textit{Danio rerio}) (35, 36), 24 compounds for European flounder (\textit{Platichthys flesus}) (35, 37, 38), 28 compounds for sheephead minnow (\textit{Cyniponodon variegates}) (39), and 34 compounds for carp (\textit{Ciprinus carpio}) (40, 41)]; however, the number of experimental data points and diversity of chemicals was not considered to be sufficient at this time for deriving meaningful correlations that would be capable of predicting the toxicity of organic compounds to these fish species.

Our method of correlation is based on the general LFER

\[
\text{SP} = c + e \cdot E + s \cdot S + a \cdot A + b \cdot B + v \cdot V
\]

(4)

where SP is the dependent variable such as the logarithm of the water to organic solvent partition coefficient or, as in the present case, the negative logarithm of the median lethal molar concentration. The independent variables (\( E, S, A, B, \) and \( V \)) are solute properties or descriptors that describe the various types of interactions involving the solute and its neighboring environment. The remaining quantities (\( c, e, s, a, b, \) and \( v \)) represent process or equation coefficients. The numerical values of the equation coefficients will be different for each aquatic organism.

Molecular descriptors for all of the compounds considered in the present study are tabulated in Table 8 of the Supporting Information. The tabulated values of a few compounds may differ slightly from values in earlier publications. The numerical values are periodically updated as additional experimental data become available. Numerical values are listed for both of the hydrogen bond basicity descriptors, \( B \) and \( B^* \). The normal hydrogen bond basicity descriptor, \( B \), is required in most of our published correlations. The alternative hydrogen bond basicity descriptor, \( B^* \), is used for select solutes in water–solvent systems when the “wet” solvent contains appreciable quantities of water, as might be the case in regards to the aquatic organisms. For most solutes, \( B \) and \( B^* \) are numerically equal but do differ mainly for alkylanilines, alklypyridines, and sulfoxides. The numerical values in Table 8 (Supporting Information) came, for the most part, from our solute descriptor database, which now contains values for more than 3500 different organic and organometallic compounds. For compounds not in our database, the descriptor values were calculated in accordance with our published computational methodology (42–47). The characteristic McGowan volume, \( V \), is calculated from the individual atomic sizes and numbers of bonds in the molecule (42). For liquid solutes, the excess molar refraction descriptor, \( E \), is obtained from the liquid refractive index (43). In the case of solid solutes, one either estimates a hypothetical liquid refractive index using any of several available methods or can calculate \( E \) directly through addition of fragments or substructures. Numerical values of the three remaining descriptors, \( S, A, \) and \( B \) (or \( B^* \)), are determined through regression analysis using available organic solvent/water partition coefficient, chromatographic retention data, solubilities, and infinite dilution activity coefficients as described elsewhere (44–47). If one is unable to find sufficient experimental data for performing the fore-mentioned regression analysis, commercial software (48) is available for estimating the molecular solute descriptors from the structure of the compound.

### Results

Our search of the chemical literature found the 96 h median lethal molar concentration of 215 fairly common organic chemicals, most of which would be expected to exhibit either a nonpolar or a polar narcotic mode of toxic action to fathead minnow. The experimental data are tabulated in Table 1 (Supporting Information). A few known electrophilic and nucleophile toxicants are deliberately inserted into the data set so that we could have some idea of the excess toxicity that nonnarcotic compounds might display. Our initial regression analyses of the data set in accord to eq 4 correctly identified the known electrophilic and nucleophile toxicants that had been included in the data set. The initial analyses indicated that 1,3-dinitrobenzene, 1,4-dinitrobenzene, 2-chlorophenol, catechol, resorcinol, pyridine, 2-chloroaniline, 2-methylimidazole, caffeine and acrolein were outliers, suggesting that their mode of action involved some type of chemical specific toxicity. These observations are in accord with the earlier observations of Ramos et al. (12) and Gunatilleka and Poole (18). The 10 outliers were removed from the data set.

The remaining quantities (\( c, e, s, a, b, \) and \( v \)) represent process or equation coefficients. The numerical values of the equation coefficients will be different for each aquatic organism.

Molecular descriptors for all of the compounds considered in the present study are tabulated in Table 8 of the Supporting Information. The tabulated values of a few compounds may differ slightly from values in earlier publications. The numerical values are periodically updated as additional experimental data become available. Numerical values are listed for both of the hydrogen bond basicity descriptors, \( B \) and \( B^* \). The normal hydrogen bond basicity descriptor, \( B \), is required in most of our published correlations. The alternative hydrogen bond basicity descriptor, \( B^* \), is used for select solutes in water–solvent systems when the “wet” solvent contains appreciable quantities of water, as might be the case in regards to the aquatic organisms. For most solutes, \( B \) and \( B^* \) are numerically equal but do differ mainly for alkylanilines, alklypyridines, and sulfoxides. The numerical values in Table 8 (Supporting Information) came, for the most part, from our solute descriptor database, which now contains values for more than 3500 different organic and organometallic compounds. For compounds not in our database, the descriptor values were calculated in accordance with our published computational methodology (42–47). The characteristic McGowan volume, \( V \), is calculated from the individual atomic sizes and numbers of bonds in the molecule (42). For liquid solutes, the excess molar refraction descriptor, \( E \), is obtained from the liquid refractive index (43). In the case of solid solutes, one either estimates a hypothetical liquid refractive index using any of several available methods or can calculate \( E \) directly through addition of fragments or substructures. Numerical values of the three remaining descriptors, \( S, A, \) and \( B \) (or \( B^* \)), are determined through regression analysis using available organic solvent/water partition coefficient, chromatographic retention data, solubilities, and infinite dilution activity coefficients as described elsewhere (44–47). If one is unable to find sufficient experimental data for performing the fore-mentioned regression analysis, commercial software (48) is available for estimating the molecular solute descriptors from the structure of the compound.

### Table 1. Intercorrelation Matrix

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>S</th>
<th>A</th>
<th>B</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>0.524</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.039</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.042</td>
<td>0.059</td>
<td>0.043</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>0.034</td>
<td>0.048</td>
<td>0.002</td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>0.001</td>
<td>0.049</td>
<td>0.013</td>
<td>0.038</td>
<td>0.125</td>
</tr>
</tbody>
</table>

where \( N = 196, R^2 = 0.953, SD = 0.276, \) and \( F = 779.4 \); the standard deviations of the coefficients are in parentheses. The correlation matrix, in \( R^2 \), between the descriptors is given in Table 1. Intercorrelations between most of the descriptors are negligible, and even the largest intercorrelation between \( E \) and \( S, 0.524 \), is not very significant. We include the descriptor, \( I \), because we have used this in other correlations. All regression analyses were performed using Minitab software (49). Equation 5 uses the \( B \) solute hydrogen bond basicity descriptor. We elected to use the \( B \) descriptors (rather than the \( B^* \) descriptors) in all of our derived toxicity correlations because readers are more likely to be able to find \( B \) values in the descriptor tabulations contained in our previous publications. We have given in Table 8 of the Supporting Information numerical values of \( B^* \) in case readers wish to use the correlations reported by Gunatilleka and Poole (18, 19) to predict the toxicity of organic chemicals to \( D. magna, Artemina salina, Pseudomonas putida, \) and \( Scendesmus quadricauda. \)

The 196 compounds were divided into a training set and a test set by ordering the compounds in terms of
increasing value of $-\log LC_{50}$. Every second compound was removed from the list to form the test set. The remaining 98 compounds that were left served as the training set. Analyses of the experimental data in the training set gave

$$\log LC_{50} = 0.926(0.102) + 0.407(0.120)E - 0.118(0.128)S + 0.304(0.124)A - 3.452(0.171)B + 3.401(0.097)V \quad (6)$$

where $N = 98$, $R^2 = 0.955$, $SD = 0.278$, and $F = 396.1$. The training set was then used to predict $-\log LC_{50}$ values for the remaining 98 compounds in the test set, to assess the correlation’s predictive ability. For the predicted and experimental values, we find that $SD = 0.285$, $AAE$ (average absolute error) = 0.226, and $AE$ (average error) = −0.054. There is therefore virtually no bias in the predictions using eq 5 with $AE$ equal to 0.054 log units.

As part of the present study, we wanted to determine whether it was possible to improve the correlation by introducing an indicator variable for the ester compounds. As noted previously, numerous esters do exhibit higher toxicities than expected of polar narcotic compounds. There are enough esters in the fathead minnow database to perform this analysis. The indicator variable $I$ is set to 1 for a compound containing an ester functional group (also set to 1 for dibutyl phosphate) and to 0 otherwise. All esters were now included in the computations. Regression analyses showed

$$-\log LC_{50} = 1.147(0.073) + 0.433(0.084)E - 0.234(0.091)S + 0.469(0.088)A - 3.648(0.122)B + 3.269(0.068)V + 0.554(0.078)I \quad (7)$$

(where $N = 205$, $R^2 = 0.951$, $SD = 0.282$, and $F = 641.3$) that the $I$ variable did improve the calculations, particularly for those esters that were removed from the eq 5 determination. For the coefficient of $I$, the $p$ value is $<0.0005$ and the $t$-test is 7.06, so the coefficient is highly significant. Without the $I$ term, the standard deviation increased to $SD = 0.297$. The improved descriptive ability is not that apparent in the statistical information because of the large number of nonester compounds in the database. For the fathead minnow, the nonester compounds dominate the regression analyses. Rather the improved descriptive ability is more noticeable in the actual predicted values for the linear alkyl esters [$-\log LC_{50} = 1.91$ (with $I$) vs $-\log LC_{50} = 1.47$ (without) vs $-\log LC_{50} = 2.32$ (exp) for methyl acetate, etc.] As an example of goodness of fit, we give in Figure 1 a plot of the calculated values on eq 5 against the observed values of $-\log LC_{50}$.

The guppy toxicity database is the second largest of the six fish species considered in the present study. Here, we were able to retrieve $-\log LC_{50}$ values for 148 compounds (see Table 2 of the Supporting Information) from the published literature for which the solute descriptors were known. None of the compounds were eliminated from the final derived correlation

$$-\log LC_{50} = 0.811(0.111) + 0.782(0.130)E - 0.230(0.136)S + 0.341(0.103)A - 3.050(0.144)B + 3.250(0.099)V \quad (8)$$

where $N = 148$, $R^2 = 0.946$, $SD = 0.280$, and $F = 493.1$.

The statistics of correlation are quite good given the nature of the property being correlated. Generally, biological data have greater experimental uncertainties associated with the reported values than do chemical properties such as the octanol/water partition coefficient or saturation solubility. Although we found independent replicate measurements for relatively few of the compounds studied, we believe that an uncertainty of ±0.20–0.25 log units (perhaps even slightly larger) would not be an unreasonable estimate for many of the experimental values in Tables 1–7 of the Supporting Information. This estimate is based on the independent measurements of Juhnke and Lüdemann (50) for the golden orfe. Figure 2 shows a graphical comparison of the calculated values based on eq 8 and the observed $-\log LC_{50}$ values for guppy. As before, the guppy toxicity database was divided into a 74 compound training set and 74 compound test set based on $-\log LC_{50}$ numerical values. Analyses of the experimental data in the training set gave

$$-\log LC_{50} = 0.697(0.135) + 0.700(0.167)E - 0.213(0.174)S + 0.429(0.144)A - 3.309(0.183)B + 3.446(0.133)V \quad (9)$$

where $N = 74$, $R^2 = 0.956$, $SD = 0.262$, and $F = 295.0$. The training set was then used to predict $-\log LC_{50}$ values for the remaining 74 compounds in the test set, to assess the correlation’s predictive ability. For the predicted and experimental values, we find that $SD = 0.316$, $AAE = 0.250$, and $AE = 0.036$. There is therefore virtually no bias in the predictions using eq 8 with $AE$ equal to 0.036 log units.

In Table 3 of the Supporting Information are listed values of the logarithm of the 96 h median lethal molar concentration for 69 organic chemicals to bluegill. Our
initial regression analysis on the data set indicated that 1,2,3,5-tetrachlorobenzene, 2-chlorophenol, and dimethyl phthalate were outliers. The measured $-\log LC_{50}$ value of 1,2,3,5-tetrachlorobenzene is too small and is not consistent with the experimental values of the other polychlorinated benzene derivatives in this data set. 2-Chlorophenol and dimethyl phthalate both showed excess toxicity in the preliminary correlation, similar to compounds previously classified as exhibiting specific toxicity. 2-Chlorophenol was an outlier in the fathead minnow correlation, and dimethyl phthalate contains two ester functional groups. The mode of toxic action of the compounds previously classified as exhibiting specific toxicity. 2-Chlorophenol and dimethyl phthalate both showed noticeable excess toxicities. The former group includes diethyl phthalate, diisobutyl phthalate, methyl 4-nitrobenzoate, and dibutyl phthalate. The Abraham model without indicator variable did improve the correlation. For the $I$ variable, $p = 0.013$ and the $t$-test is 2.71, so the variable is very significant. Without the $I$ variable, the standard deviation for the 59 compounds increased to $SD = 0.333$. The improvement descriptive ability is more apparent in the statistical information for the golden orfe database because the ester toxicients comprise a greater percentage of the golden orfe toxicity database. In the case of fathead minnow, there were too few esters in the toxicity database for the improved descriptive ability to be noticeable in the statistical information.

On the basis of our computations to date and our reading of the published literature, we have divided the esters in this study into two groups, those that exhibit negligible or small excess toxicities vs those that exhibit fairly significant excess toxicities. The former group includes diethyl phthalate, diisobutyl phthalate, methyl 4-chlorobenzoate, ethyl 4-aminobenzoate, and dibutyl phthalate. The Abraham model without indicator variable correlates the measured $-\log LC_{50}$ values of these esters, along with the measured $-\log LC_{50}$ values of the other nonpolar and polar narcotic toxicients. Most of the linear alky esters, ethyl benzoate, diphenyl phthalate, and methyl 4-nitrobenzoate, fall in the latter group. The indicator variable is needed to improve the aquatic toxicity correlation. There are of course several esters that fall near the dividing line. Either form of the Abraham model works equally well for such esters. Our calculations on the fathead minnow and golden orfe databases suggest that this grouping may be transferable between the different species of fish.

**Discussion**

Examination of the numerical entries in Table 2 reveals that the correlation coefficients do vary from one fish species to another. For some species of fish, the differences are quite significant. At the moment, we have no good explanation for why the $c$ and $s$ coefficients are so different for the 48 vs 96 h medaka high-eyes correlations. The effect may be real or may simply be an artifact associated with one (or both) of the data sets used in the regression analysis. Derived correlations do depend on...
the accuracy of the experimental data used and on the range of molecular descriptors spanned by the data set(s). Several bad data points will affect the numerical values of the derived coefficients more in a smaller data set than in a larger data set. Also, smaller data sets are more likely to have a more dissimilar distribution of nonpolar vs polar and of acidic vs basic compounds. These are problems commonly encountered in developing and comparing any quantitative structure–activity/property relationship (QSAR/QSPR) or LFER. We do note that it is possible to assess the consistency of a small portion of the experimental medaka high-eyes data in that the 98 h median lethal molar concentration must be less than or equal to the 48 h value, i.e., −log LC_{50,96h} ≥ −log LC_{50,48h}. Careful examination of Tables 4 and 5 (Supporting Information) reveals that data for 1,2-dichlorobenzene, benzene, toluene, and 4-chloroaniline do not satisfy this requirement. To determine whether these four compounds were responsible for the large differences observed in the equation coefficients, the four compounds were removed from both the 96 and the 48 h medaka high-eyes data sets, and the regression analysis was performed again. Removal of the four compounds had only a small effect on the numerical values of the equation coefficients.

One of the advantages of using a particular equation, such as the Abraham equation, to analyze data in different systems, is that considerable information can be derived from the coefficients in the equation. These coefficients are specific for a particular system and encode chemical information about the system. Thus, partitions in the water–octanol system can be correlated (31) through the general LFER, eq 4, to yield

\[
\log P_{\text{oct}} = 0.088 + 0.562E - 1.054S + 0.034A - 3.460B + 3.814V
\]

Table 2. Correlations for the Nonspecific Aquatic Toxicity of Organic Compounds to Different Species of Fish

<table>
<thead>
<tr>
<th>system</th>
<th>no.</th>
<th>c</th>
<th>e</th>
<th>s</th>
<th>a</th>
<th>b</th>
<th>v</th>
<th>( R^2 )</th>
<th>SD</th>
<th>F</th>
<th>N</th>
<th>SD_{pred}</th>
</tr>
</thead>
<tbody>
<tr>
<td>fathead minnow</td>
<td>0.996</td>
<td>0.418</td>
<td>−0.182</td>
<td>0.417</td>
<td>−3.574</td>
<td>3.377</td>
<td>0.953</td>
<td>0.276</td>
<td>779.5</td>
<td>196</td>
<td>0.285</td>
<td></td>
</tr>
<tr>
<td>P. promelas</td>
<td>0.811</td>
<td>0.782</td>
<td>−0.230</td>
<td>0.341</td>
<td>−3.050</td>
<td>3.250</td>
<td>0.946</td>
<td>0.280</td>
<td>493.1</td>
<td>148</td>
<td>0.316</td>
<td></td>
</tr>
<tr>
<td>guppy (P. reticulata)</td>
<td>0.903</td>
<td>0.583</td>
<td>−0.127</td>
<td>1.238</td>
<td>−3.918</td>
<td>3.306</td>
<td>0.968</td>
<td>0.272</td>
<td>359.8</td>
<td>66</td>
<td>0.264</td>
<td></td>
</tr>
<tr>
<td>bluegill (L. macrochirus)</td>
<td>0.877</td>
<td>0.499</td>
<td>0.000</td>
<td>1.223</td>
<td>−4.014</td>
<td>3.322</td>
<td>0.967</td>
<td>0.271</td>
<td>452.9</td>
<td>66</td>
<td>0.264</td>
<td></td>
</tr>
<tr>
<td>goldfish (C. auratus)</td>
<td>0.922</td>
<td>−0.653</td>
<td>1.872</td>
<td>−3.292</td>
<td>3.244</td>
<td>0.935</td>
<td>0.269</td>
<td>127.0</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>olive oil</td>
<td>0.834</td>
<td>1.047</td>
<td>−0.380</td>
<td>0.806</td>
<td>−4.516</td>
<td>3.078</td>
<td>0.966</td>
<td>0.277</td>
<td>253.7</td>
<td>51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medaka high-eyes (O. latipes) 48-hr. end point</td>
<td>0.802</td>
<td>−1.046</td>
<td>0.272</td>
<td>0.931</td>
<td>−2.178</td>
<td>3.155</td>
<td>0.960</td>
<td>0.277</td>
<td>181.8</td>
<td>44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>medaka high-eyes (O. latipes) 96-hr. end point</td>
<td>0.802</td>
<td>1.046</td>
<td>0.272</td>
<td>0.931</td>
<td>−2.178</td>
<td>3.155</td>
<td>0.960</td>
<td>0.277</td>
<td>181.8</td>
<td>44</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( a \) Standard deviation for “test” data set predictions using the correlation equation derived from “training” data set.

Table 3. Equation Coefficients for Various Organic Solvent–Water Partitions and for the Median Lethal Toxicities to Various Species of Fish

<table>
<thead>
<tr>
<th>system</th>
<th>no.</th>
<th>e</th>
<th>s</th>
<th>a</th>
<th>B</th>
<th>V</th>
<th>ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>octanol</td>
<td>1</td>
<td>0.562</td>
<td>−1.054</td>
<td>0.034</td>
<td>−3.460</td>
<td>3.814</td>
<td>51</td>
</tr>
<tr>
<td>isobutanol</td>
<td>2</td>
<td>0.514</td>
<td>−0.693</td>
<td>0.020</td>
<td>−2.255</td>
<td>2.776</td>
<td>55</td>
</tr>
<tr>
<td>pentanol</td>
<td>3</td>
<td>0.575</td>
<td>−0.787</td>
<td>0.020</td>
<td>−2.837</td>
<td>3.249</td>
<td>62</td>
</tr>
<tr>
<td>4-ethyl alcohol</td>
<td>4</td>
<td>−0.270</td>
<td>−0.528</td>
<td>−0.035</td>
<td>−4.042</td>
<td>4.204</td>
<td>55</td>
</tr>
<tr>
<td>dichloromethane</td>
<td>5</td>
<td>0.001</td>
<td>0.022</td>
<td>−3.238</td>
<td>−4.137</td>
<td>4.259</td>
<td>45</td>
</tr>
<tr>
<td>trichloromethane</td>
<td>6</td>
<td>0.157</td>
<td>−0.391</td>
<td>−3.191</td>
<td>−3.437</td>
<td>4.191</td>
<td>54</td>
</tr>
<tr>
<td>tetrachloromethane</td>
<td>7</td>
<td>0.573</td>
<td>−1.254</td>
<td>−3.558</td>
<td>−4.588</td>
<td>4.599</td>
<td>45</td>
</tr>
<tr>
<td>diethyl ether</td>
<td>8</td>
<td>0.561</td>
<td>−1.016</td>
<td>−0.226</td>
<td>−4.553</td>
<td>4.075</td>
<td>62</td>
</tr>
<tr>
<td>dibutyl ether</td>
<td>9</td>
<td>0.677</td>
<td>−1.506</td>
<td>−0.807</td>
<td>−5.249</td>
<td>4.815</td>
<td>63</td>
</tr>
<tr>
<td>NPOE</td>
<td>10</td>
<td>0.600</td>
<td>−0.459</td>
<td>−2.246</td>
<td>−3.879</td>
<td>3.574</td>
<td>56</td>
</tr>
<tr>
<td>ethyl acetate</td>
<td>11</td>
<td>1.157</td>
<td>−1.397</td>
<td>−0.054</td>
<td>−3.755</td>
<td>3.726</td>
<td>55</td>
</tr>
<tr>
<td>PGDPc</td>
<td>12</td>
<td>0.501</td>
<td>−0.828</td>
<td>−1.022</td>
<td>−4.640</td>
<td>4.033</td>
<td>55</td>
</tr>
<tr>
<td>olive oil</td>
<td>13</td>
<td>0.574</td>
<td>−0.798</td>
<td>−1.422</td>
<td>−4.884</td>
<td>4.210</td>
<td>55</td>
</tr>
<tr>
<td>benzene</td>
<td>14</td>
<td>0.464</td>
<td>−0.588</td>
<td>−3.099</td>
<td>−4.625</td>
<td>4.491</td>
<td>45</td>
</tr>
<tr>
<td>nitrobenzene</td>
<td>15</td>
<td>0.576</td>
<td>0.003</td>
<td>−2.356</td>
<td>−4.420</td>
<td>2.634</td>
<td>64</td>
</tr>
<tr>
<td>hexane</td>
<td>16</td>
<td>0.579</td>
<td>−1.723</td>
<td>−3.599</td>
<td>−4.764</td>
<td>4.344</td>
<td>45</td>
</tr>
<tr>
<td>hexadecane</td>
<td>17</td>
<td>0.667</td>
<td>−1.617</td>
<td>−3.587</td>
<td>−4.869</td>
<td>4.433</td>
<td>45</td>
</tr>
<tr>
<td>cyclohexane</td>
<td>18</td>
<td>0.784</td>
<td>−1.678</td>
<td>−3.740</td>
<td>−4.929</td>
<td>4.577</td>
<td>65</td>
</tr>
<tr>
<td>carbon disulfide</td>
<td>19</td>
<td>0.886</td>
<td>−0.943</td>
<td>−3.603</td>
<td>−5.818</td>
<td>4.931</td>
<td>53</td>
</tr>
<tr>
<td>fathead minnow, eq 5</td>
<td>20</td>
<td>0.418</td>
<td>−0.182</td>
<td>0.417</td>
<td>−3.574</td>
<td>3.377</td>
<td>this work</td>
</tr>
<tr>
<td>guppy, eq 8</td>
<td>21</td>
<td>0.782</td>
<td>−0.230</td>
<td>0.341</td>
<td>−3.050</td>
<td>3.250</td>
<td>this work</td>
</tr>
<tr>
<td>bluegill, eq 10</td>
<td>22</td>
<td>0.583</td>
<td>−0.127</td>
<td>1.238</td>
<td>−3.918</td>
<td>3.306</td>
<td>this work</td>
</tr>
<tr>
<td>golden orfe</td>
<td>23</td>
<td>0.931</td>
<td>0.379</td>
<td>0.951</td>
<td>−2.392</td>
<td>3.244</td>
<td>0.935</td>
</tr>
<tr>
<td>goldfish</td>
<td>24</td>
<td>−0.653</td>
<td>1.872</td>
<td>−0.329</td>
<td>−4.516</td>
<td>4.078</td>
<td>0.938</td>
</tr>
<tr>
<td>medaka, 48 h</td>
<td>25</td>
<td>1.047</td>
<td>−0.380</td>
<td>0.806</td>
<td>−2.182</td>
<td>2.667</td>
<td>452.9</td>
</tr>
<tr>
<td>medaka, 96 h</td>
<td>26</td>
<td>1.046</td>
<td>0.272</td>
<td>0.931</td>
<td>−2.178</td>
<td>3.155</td>
<td>0.960</td>
</tr>
</tbody>
</table>

\( a \) PGDP is propylene glycol dipelargonate.
It can then be seen that octanol is much more hydrophobic than water (v is positive) but is a much weaker hydrogen bond acid (b is negative). What is surprising, but follows inexorably from eq 4, is that octanol and water have almost the same hydrogen bond acidity (a is nearly zero). Not only can this type of analysis be carried out for any system, but comparison of coefficients between systems will indicate how near one system is to another, in terms of chemical interactions, and hence whether one system can be used as model for another.

A number of water–solvent systems have been suggested as models for biological systems, starting with the use of water–olive oil by Meyer (52) and by Overton (29, 30). Other systems were subsequently studied in the 1930s, including water–oleyl alcohol, water–benzene, water–cyclohexane, and water–carbon disulfide (53). Water–chloroform (54), water–isopropyl myristate (IPM) (55), and water–o-nitrophenyl octyl ether (56) have all been used, but through the work of Hansch and Leo, the water–octanol system has become the system of choice (57, 58). We collect in Table 3 the coefficients (e, s, a, b, and v) for the systems that we have studied and for a variety of water–solvent partitioning systems. If only a few systems are to be compared, the coefficients can be examined by eye, but for a large number of systems, some form of data analysis is essential. Two mathematical procedures are those of Abraham and Martins (59, 60), who calculate the five-dimensional distance between the coefficients as points in five-dimensional space, and of Ishihama and Asakawa (67), who calculate the angle between the coefficients now regarded as lines in five-dimensional space. A method that is visually more accessible is simply to carry out a principal components analysis (PCA) on the five columns of coefficients, thus converting them into five PCs. The first two PCs contain most of the information, and when the scores of PC1 and PC2 are plotted against each other, the chemical closeness of any two systems is reflected in how close the corresponding points are in two-dimensional space. In the PCA, we used the correlation matrix in which the variables are scaled. However, because our variables cover similar ranges, it makes little difference if we use the covariance matrix in which the variables are not scaled.

The PC score plot is shown in Figure 4. Note that the point for goldfish (no. 24) is so far away from all of the other points that it is just indicated on the plot. Except for goldfish (no. 24), the biological systems fall into two quite close units, one containing fathead minnow (no. 20), the guppy (no. 21), and bluegill (no. 22), and the other containing the golden orfe (no. 23) and the two medaka equations (nos. 25 and 26). Of the water–solvent systems, water–isobutanol (no. 2) is squarely in the middle of the two groups, and water–pentanol (no. 3) is in the second group. It is noteworthy that the organic phase in these two cases contains a considerable amount of water, even more so than water-saturated octanol (no. 1), which is some way away from the biological systems. We can conclude that the best chemical model for nonspecific aquatic toxicity in general is water–isobutanol but that water–pentanol is very close to systems nos. 23, 25, and 26. The water–octanol system and the remaining water–solvent systems in Table 3 are all poor or very poor chemical models.

We have no explanation as to why the goldfish equation is so far away from all of the others. The statistics of the equation, Table 2, are not exceptional, and there is a reasonable number of data points, 51, in the correlation. Clearly, even if a suitable model is found for one species of fish, it cannot be assumed that the same model will be appropriate for another fish species.

Acknowledgment. This research was supported in part by the University of North Texas Research Council. K.H. thanks the University of North Texas and the U.S. Department of Education for support provided under the Ronald E. McNair Postbaccalaureate Achievement Program.

Supporting Information Available: Median lethal molar concentration toxicity data and table of solute descriptors. This material is available free of charge via the Internet at http://pubs.acs.org.

References


Figure 4. Plot of the scores for PC2 against the scores for PC1. Points numbered as in Table 3.


TX050164Z