Experimental Determination of Octanol-Water Partition Coefficients of Selected Natural Toxins

Carina D. Schönsee, Thomas D. Bucheli

Environmental Analytics, Agroscope, Reckenholzstrasse 191, 8046 Zürich, Switzerland

* Corresponding author

Telephone number: 0041 44 377 73 42

E-mail address: thomas.bucheli@agroscope.admin.ch
Abstract (200 words maximum)

Natural toxins are ubiquitously occurring highly diverse organic compounds produced by e.g., plants or fungi. In predictive environmental fate and risk assessment of organic chemicals for regulatory purposes, the octanol-water partition coefficient ($K_{ow}$) remains one of the key parameters. However, experimental data for natural toxins is largely missing and current estimation models for $K_{ow}$ show limited applicability for multifunctional, ionizable compounds. Thus, log $K_{ow}$ data was first experimentally derived for a diverse set of 45 largely ionizable natural toxins and then compared to predicted values from three different models (KOWWIN, ACD/Percepta, Chemicalize). Both approaches were critically evaluated with regards to their applicability for multifunctional, ionizable compounds. The miniaturized shake-flask approach allowed reliable quantification of pH dependent partitioning behavior for neutral, acidic and basic ionizable natural toxins. All analyzed toxins are rather polar with an average log $K_{ow} < 1$ and an observed maximum log $K_{ow}$ of 2.7. Furthermore, the comparison of experimental data to those of commonly used prediction models showed that the latter match the former with only minorly increased errors. The Chemicalize tool gave overall best predictions with a mean absolute error of 0.49 and thus should be preferred in comparison to ACD/Percepta and KOWWIN.
Introduction

The number of investigations on environmental behavior of synthetic chemicals (e.g., pesticides) is continuously increasing.\(^1\)\(^-\)\(^2\) In contrast, a huge knowledge gap exists when it comes to the occurrence and distribution of naturally toxic compounds. Studies on natural toxins are limited to just a few compounds found in water resources.\(^3\)\(^-\)\(^7\) Mycotoxins and plant secondary metabolites (phytotoxins) are two of the largest subgroups of natural toxins. For their producers they act as advantageous protection or defense molecules against herbivores, microbes, viruses or other plants.\(^8\)\(^,\)\(^9\) As such they can be seen as nature’s own pesticides that may pose a threat to husbandry animal and human health though.\(^10\)\(^-\)\(^12\) Their molecular diversity is immense, spanning from alkaloids to steroids, flavonoids, terpenoids and many others.\(^13\)

Once released to the environment, natural toxins, not unlike anthropogenic pollutants, are subject to different fate processes that are largely dependent on the compounds’ physicochemical properties. For the purpose of environmental risk assessment and as basis for predictive modeling and remediation strategies, those properties affecting a compound’s distribution have to be described in a systematic and quantitative manner.\(^14\) The octanol-water partition coefficient (\(K_{ow}\)) remains one of the key parameters in environmental fate and risk assessment studies of organic chemicals. Many single-parameter Quantitative Structure-Property (QSPRs) and Activity Relationships (QSARs) rely on \(K_{ow}\) as main input parameter.\(^15\)\(^-\)\(^18\) In bioconcentration and toxicity estimation in particular, \(K_{ow}\) continues to be an important estimate that allows to derive a tendency on hydrophobicity or lipophilicity of a compound.\(^16\)\(^,\)\(^19\)

The \(K_{ow}\) is only valid for a neutral chemical, while partitioning of ionizable chemicals is described by the respective distribution coefficient (\(D_{ow}\)), accounting for both neutral and charged species of the chemical. In the environmentally relevant pH range of pH 4 to 9, \(D_{ow}\) is considered more relevant for the description of environmental behavior of ionizable compounds.\(^15\)

Different well-established methods are available for the experimental determination of \(K_{ow}/D_{ow}\) (e.g., OECD 107, 117).\(^20\)\(^,\)\(^21\) The indirect analysis of partitioning behavior based on column chromatography (i.e., OECD 117)\(^21\) allows simple and fast analysis of larger groups of
compounds by pooled injection. However, it was shown that data reproducibility is rather poor and that charge as well as steric effects may outcompete hydrophobic effects in retention analysis. Thus, the shake-flask technique (OECD 107) is currently preferred for reliable determination of $K_{ow}/D_{ow}$, particularly when focusing on ionizable compounds with an expected log $K_{ow}/D_{ow} < 4.5$. The method can be applied to ionizable compounds by performing experiments in dependence of pH. Miniaturization can speed up total measurement time and considerably minimizes required amounts of analytes and laboratory consumables. This was demonstrated by Rothwell et al. for dietary flavonoids. In combination with a modified in situ approach developed for application in drug analysis, the approach can be seen as potential high-throughput alternative for analysis of ionizable organics.

Experimental $K_{ow}/D_{ow}$ data for natural toxins is extremely scarce. Hence, current risk assessment often has to resort to in silico prediction tools. However, for compounds with physicochemical and structural complexity due to large numbers of ionizable functional groups, current estimation models for phase distribution coefficients show limited applicability.

Thus, reliable experimental data is required to improve understanding of natural toxin behavior in the environment. In this study, we provide experimentally derived log $K_{ow}/D_{ow}$ values for a set of 45 mainly ionizable natural toxins from different compound classes prioritized as potential aquatic micropollutants. In addition, we critically evaluate the applicability of both experimental (miniaturized shake-flask approach) and predictive methods (EPiSuite, ACD/Percepta, Chemicalize) for $K_{ow}/D_{ow}$ determination.

Materials and Methods

Materials

The chemicals used in this study were purchased from the following companies: HPLC grade n-octanol and methanol as well as sodium salts used for preparation of buffer solutions from Sigma-Aldrich (Buchs, Switzerland); investigated natural toxins from PhytoLab (Vestenbergsreuth, Germany), Sigma-Aldrich (Buchs, Switzerland), or Fermentek (Jerusalem, Israel). Water was deionized using a Milli-Q system (Merck Millipore, Darmstadt, Germany).
All natural toxin stock solutions were prepared in methanol. Details on analytes (CAS, formula, molecular structure), suppliers and stock solution concentrations are given in the supporting information (SI, Table SI-1 and 4, Figure SI-1). The chosen set of 45 natural toxins (37 phytotoxins, 8 mycotoxins, Table 1) covered a molecular weight range of 150 to 650 g mol\(^{-1}\) and contained representatives of neutral, charged, acidic as well as basic ionizable compounds.

**Experimental Approach**

Octanol-water partition coefficients were quantified using a miniaturized shake-flask approach based on OECD 107 (Figure SI-2).\(^{19, 20, 30}\) To investigate analytes in either fully ionized or neutral state as well as their pH dependent partitioning behavior, measurements were performed using sodium based buffer solutions at a constant ionic strength of 0.1 M covering three different pH: pH 4.0 (citrate buffer), pH 7.3 (phosphate buffer) and pH 10.2 (carbonate buffer). All buffer solutions were saturated with octanol prior to analysis and vice versa. Stock solutions of natural toxins were first diluted 1:100 in defined volumes of buffer solution and added to standard 1.5 mL HPLC vials. Secondly, octanol was added to all but one of the dilutions which served as reference stock dilution. Varying ratios of octanol (0.01-1 mL) and aqueous buffer solutions (0.5-1.5 mL) were chosen based on the predicted \(K_{\text{ow}}\), with the goal that between 30 - 80 % of the toxin would remain in the aqueous phase after partitioning. Thus, the theoretical operational range of the method was set from \(\log D_{\text{ow}}\) of -1 to 2.5 and was limited by: 1) detection limits of the DAD, 2) minimum aqueous phase volumes of 0.5 mL required for phase separation and HPLC injection. Further details on preparation of buffer solutions and octanol-water phase ratios for individual natural toxins are presented in Section II of the SI (Table SI-3 and 4).

All octanol-water mixtures were first vortexed for 1 min and then shaken for 2 h (125 rpm) at room temperature (25 ± 3°C) to reach equilibrium and phase distribution. Kinetic measurements were performed to confirm equilibrium conditions in the setup (data not shown). After equilibration and phase separation, the aqueous phase of all samples was analyzed on an Agilent HPLC 1260 with DAD detection (\(\lambda = 210, 228 \text{ or } 248 \text{ nm, analyze dependent, details} \))
Applying the water plug technique, 50 µl of the aqueous phase were injected. In brief, first 5 µl of blank buffer solution were aspirated, followed by the aqueous phase of the respective sample. To avoid any carryover of the analyte containing octanol phase, the outside of the syringe was subsequently washed in methanol before injection. Analysis was done with a Macherey Nagel Nucleoshell RP 18plus column (length 50 mm, ID 2 mm, particle size 2.7 µm) and without any column for comparison. Runs were performed at 40°C in isocratic mode using a methanol/water mix (40/60, v/v) at a flow rate of 0.33 mL min\(^{-1}\). As resulting partition coefficients did not significantly differ between analysis with and without column, final results were obtained from measurements without a column. This led to a substantial decrease in analysis time and prevented potential memory effects due to accumulation of octanol on the stationary phase of the column. However, obtained peak areas for analytes had to be corrected with peak areas of the aqueous (octanol-saturated) buffer solution blanks at the respective measurement wavelength.

**Data Evaluation**

Distribution coefficients for all analytes were derived based on the difference of the blank corrected peak areas of the toxin in both the reference stock dilution (area\(_{\text{Std}}\)) and the aqueous phase of the partition samples containing octanol (area\(_{w}\)), multiplied with the volumetric ratio of water (V\(_{\text{water}}\)) and octanol (V\(_{\text{octanol}}\)) of the respective sample (given in Table SI-4) according to Equation 1.

\[
\log D_{ow} [-] = \log \left( \frac{\text{area}_{\text{Std}}}{\text{area}_{w}} - 1 \right) \times \frac{\text{V}_{\text{water}} [mL]}{\text{V}_{\text{octanol}} [mL]}
\]

Equation 1

Overall method applicability was tested using previously investigated natural toxins for which experimental log K\(_{ow}\) is available (caffeine, colchicine, daidzein; Table SI-2) as reference compounds. For comparing data to commonly used prediction models, only experimentally derived partitioning data for compounds examined in their neutral state was considered. In this case log D\(_{ow}\) = log K\(_{ow}\) and no errors due to partial partitioning of ionized species are introduced. To calculate the ionization state of a given analyte under experimental conditions, the ionized fraction was obtained based on the compound’s pK\(_a\) (ACD/Percepta\(^{39}\), Table 1) and
measurement pH (Table SI-3). The calculation is based on the relationships stated in Equation 2 (basic ionizable analytes) and Equation 3 (acidic ionizable analytes). If the ionized fraction was lower than 1%, a compound was assumed to be in its neutral state.

\[
\text{ionized fraction, base} [%] = \frac{1}{10^{-pK_a} + 1} \times 100% \quad \text{Equation 2}
\]

\[
\text{ionized fraction, acid} [%] = 1 - \frac{1}{10^{-pK_a} + 1} \times 100% \quad \text{Equation 3}
\]

Commonly used QSPR models evaluated with regards to predictability of natural toxin log \( \text{K}_{\text{ow}} \) were the free of charge KOWWIN, integrated in the Estimation Program Interface (EPI) Suite from the U.S. EPA, as well as the commercially available ACD/Percepta and Chemicalize tools. Predicted log \( \text{K}_{\text{ow}} \) data for all compounds and models is given in the SI (Table SI-2). The evaluation was based on regression analysis as well as the mean absolute error (MAE) of the models according to Equation 4:

\[
\text{MAE} = \frac{1}{n} \sum |\log K_{\text{ow,exp}} - \log K_{\text{ow, model}}| \quad \text{Equation 4}
\]

where \( n \) is the number of observations and \( K_{\text{ow,exp}} \) and \( K_{\text{ow, model}} \) the experimentally derived and QSPR predicted partition coefficients, respectively.

**Results and Discussion**

**Octanol-Water Partition Coefficients and Applicability of the Experimental Approach**

Distribution coefficients (log \( \text{D}_{\text{ow}} \)) for all investigated natural toxins are given in Table 1, sorted first by the type of the toxins’ first ionizable group, and second by alphabet. It is often neglected that in many cases the ionic species also partitions into the organic phase when extrapolating log \( \text{K}_{\text{ow}} \) from log \( \text{D}_{\text{ow}} \). Deriving log \( \text{K}_{\text{ow}} \) from measurements of a partially ionized compound could thus lead to increased errors. Therefore, only for those compounds analyzed in their neutral state (< 1% ionized), log \( \text{K}_{\text{ow}} \) values are presented in this study (Table 1, in italics).
Table 1: List of analyzed neutral, charged, basic and acidic ionizable natural toxins, including their first basic or acidic pKₐ as well as experimentally derived log Dₐw values and standard deviations at different pH and ionization states. Number in parentheses depict the number of independent measurements performed. When log Dₐw is given in italics, log Dₐw = log Kₐw as the compound was evaluated in its neutral state, defined by ionization < 1%.

| compound                        | pKₐᵃ) | fraction of charged species [%] at pH 4.0/pH 7.3/pH 10.2 | log Dₐw (pH 4.0) | log Dₐw (pH 7.3) | log Dₐw (pH 10.2) |
|---------------------------------|--------|---------------------------------------------------------|------------------|------------------|------------------|
| **neutral**                     |        |                                                         |                  |                  |                  |
| Caffeine                        | -      | neutral                                                 | -0.12 ± 0.07 (2) | -0.08 ± 0.03 (6) | -0.32 ± 0.06 (9) |
| Colchicine                      | -      | neutral                                                 | 1.11 ± 0.02 (3)  | 1.14 ± 0.03 (9)  | 1.23 ± 0.05 (9)  |
| Strophanthin                    | -      | neutral                                                 |                  | 0.69 ± 0.06 (3)  | 0.77 ± 0.05 (3)  |
| **charged**                     |        |                                                         |                  |                  |                  |
| Berberine                       | -      | cationic                                                | -0.42 ± 0.23 (3) | -1.03 ± 0.18 (3) | -0.09 ± 0.07 (3) |
| **acids**                       |        |                                                         |                  |                  |                  |
| 3-Acetyl-Deoxynivalenol         | 11.80  | -/0/-                                                   |                  |                  |                  |
| Ailanthone                      | 11.85  | 0/-/0                                                   |                  |                  |                  |
| Cucurbitacin E                  | 8.51   | -/-99                                                   |                  |                  |                  |
| 10-Deacetylbaccatin III         | 11.50  | 0/-/-                                                   |                  |                  |                  |
| Daidzein                        | 7.01   | 0/71/100                                                | 2.61 ± 0.03 (6)  | 2.22 ± 0.05 (3)  | -0.26 ± 0.13 (6) |
| Deacetoxyscirpenol              | 13.40  | -/-/0                                                   |                  |                  |                  |
| Deoxynivalenol                  | 11.90  | -/-2/-                                                  |                  |                  |                  |
| Nivalenol                       | 11.80  | -/-2/-                                                  |                  |                  |                  |
| Patulin                         | 12.10  | -/-2/-                                                  |                  |                  |                  |
| 8-Prenylnaringenin              | 7.70   | -/-100                                                  |                  |                  |                  |
| Sterigmatocystin                | 6.90   | -/-76/-                                                 |                  |                  |                  |
| β-Zearalenol                    | 7.60   | -/-39/-                                                 |                  |                  |                  |
| Zearalenone                     | 7.60   | -/-39/-                                                 |                  |                  |                  |
| **bases**                       |        |                                                         |                  |                  |                  |
| 7-Acetyllycopsamine             | 7.85   | -/-74/-                                                 |                  |                  |                  |
| Aconitin                        | 5.15   | -/-0/-                                                  |                  |                  |                  |
| (+)-Bicucullin                  | 6.71   | 100/17/-                                                | -0.49 ± 0.04 (3) | 1.86 ± 0.01 (3)  |                  |
| Cytisin                         | 10.50  | -/-66                                                   |                  |                  |                  |
| Echimidine                      | 7.36   | 100/48/0                                                | -0.93 ± 0.03 (2) | 0.64 ± 0.15 (3)  | 1.26 ± 0.14 (3)  |
| Erucifoline                     | 5.92   | 99/3/0                                                  | -0.95 ± 0.12 (3) | -0.17 ± 0.05 (3) | 0.44 ± 0.21 (3)  |
| Erucifoline N-oxide             | 4.65   | 82/-/-                                                  | -0.35 ± 0.31 (3) |                  |                  |
| Galanthamine                    | 8.49   | -/-2/-                                                  |                  |                  | 0.96 ± 0.06 (3)  |
| Gramine                         | 7.92   | -/-1/-                                                  |                  |                  | 0.88 ± 0.04 (15) |
| Huperzine A                     | 9.47   | 100/-15                                                 | -1.32 ± 0.40 (4) |                  | 1.94 ± 0.02 (3)  |
| (+)-Isocorydin                  | 6.77   | -/-0/-                                                  |                  |                  | 1.70 ± 0.02 (3)  |
| Jacobine                        | 5.86   | -/-3/-                                                  | -0.32 ± 0.38 (3) | 0.83 ± 0.24 (3)  |                  |
| Jacobine N-oxide                | 4.63   | 81/-10                                                  | -0.87 ± 0.84 (2) |                  | -1.01 ± 0.40 (3) |
| Lasicarpine                     | 7.35   | 100/47/0                                                | -0.42 ± 0.35 (3) | 1.22 ± 0.29 (3)  | 1.84 ± 0.11 (6)  |
| Lasicarpine N-oxide             | 3.87   | 43/-0/-                                                 | -0.50 ± 0.49 (3) | -0.21 ± 0.24 (6) |                  |
| Lycopsamine N-oxide             | 4.26   | 65/-10                                                  | -1.00 ± 0.43 (3) |                  | -0.57 ± 0.43 (3) |
| Lycorine                        | 6.34   | -/-8/0                                                  | -0.30 ± 0.31 (3) | 0.71 ± 0.05 (3)  |                  |
| Monocrotaline                   | 5.90   | -/-1/-                                                  |                  |                  | 0.27 ± 0.33 (6)  |
| Papaverine                      | 6.32   | 100/-/-                                                 | 0.30 ± 0.02 (3)  |                  |                  |
| Compound       | Predicted $pK_a$ | Experimental $pK_a$ | Standard Deviation | Experimental Error |
|----------------|------------------|--------------------|--------------------|--------------------|
| Protopin       | 7.86             | -0.67 ± 0.43 (4)   | -                  | -                  |
| Reserpine      | 7.25             | -0.16 ± 0.63 (2)   | -                  | -                  |
| Retrorsine     | 5.79             | -1.87 ± 0.09 (2)   | 1.03 ± 0.08 (6)    |
| Senkirkine     | 6.51             | -0.78 ± 0.01 (2)   | 0.61 ± 0.11 (3)    |
| Senecionine    | 5.86             | -0.63 ± 0.57 (3)   | 1.40 ± 0.10 (6)    |
| Seneciphylline | 5.87             | -0.91 ± 0.17 (2)   | 1.53 ± 0.02 (3)    |
| Tetrahydropalmatin | 6.53        | -0.15 ± 0.21 (3)   | -                  |
| Vincamine      | 7.82             | -0.91 ± 0.13 (3)   | -                  |

*ACD/Percepta predicted data, $pK_a$ is valid for the corresponding cation, - no data generated*

Overall results range from a minimum of log $D_{ow} = -1.87$ to a maximum of log $D_{ow} = 2.68$ with highly variable experimental errors (Table 1). The variability is a result of influences of experimental conditions such as pH and temperature stability or potential sorption of cationic analytes to glass surfaces. Ionic strength and the type of counter ions in solution are known to affect partitioning as well. All buffers were based on sodium salts and the variation of ionic strength was kept as narrow as possible (Table SI-3). However, the counter ions varied and may have increased variations of $K_{ow}$ obtained at different pH as seen for the neutral reference compounds colchicine and caffeine (Figure 1a, b). Experimental errors obtained by analysis of replicates are larger for analytes with lower affinity for octanol (Table 1, Figure SI-3). Over all results, the average standard deviation (SD) for analytes with log $D_{ow} < 0$ is 0.2 and thus twice as high as for compounds with log $D_{ow} > 0$. A small log $D_{ow}$ results in only little changes in the aqueous phase concentration that are not as easily detected as for analytes with high log $D_{ow}$. No clear tendency towards higher experimental errors with increasing degree of compound ionization is observed (Figure SI-3). However, errors are slightly elevated when the analyte is neither fully ionized nor in its neutral state under experimental conditions. The average SD for compounds with ionization between 1 and 90% is 0.24 and thus larger in comparison with neutral (SD = 0.15) and fully ionized analytes (SD = 0.13; Table 1, Figure SI-3).
Figure 1: Measured log $K_{ow}$ values for analyzed reference compounds at different pH (pH 10.2, 7.3 and 4.0) and as average of all measurements (mean). For the acidic ionizable reference daidzein (c), experimental results from measurements at pH 4.0, where the compound is in its neutral state, are shown. Including measurement errors, no statistically significant difference ($p > 0.05$) can be observed when comparing experimental values from this study and literature data (lit).^{19, 30}

Nevertheless, obtained log $K_{ow}$ values for the reference compounds are consistent (i.e., within 0.1 log units) with literature values and show good reproducibility (maximum standard deviation of the mean < 0.15, Figure 1). Thus, the miniaturized shake-flask approach provides robust data for natural toxin $K_{ow}/D_{ow}$. In the following, the data is further evaluated regarding influences of pH changes on partitioning and serves as a basis for a critical assessment of the predictive power of different QSPR models.

**Influence of pH on Octanol-Water Partitioning**

As a consequence of proton transfer reactions and formation of corresponding charged species, partitioning of ionizable compounds is pH dependent. For seven of the ionizable toxins, data for both the neutral as well as fully ionized species were acquired. An additional 16 compounds were measured only as neutral, nine as fully ionized species. All other compounds were evaluated independent of their speciation (Table 1). In the environmentally relevant range of pH 4 to 9 they would also never occur in their fully ionized or neutral form. The medium pH would need to be more than one unit above (basic ionizable) or below (acidic ionizable) the compound’s $pK_a$ to achieve a system with only one species of the compound present.^{15, 20} Additionally, measurement limitations such as limited variability of octanol-water
ratios and high pH values (pH > 10) that are not advisable for common HPLC systems did not allow deriving those values in this study.

Generally, neutral species showed higher affinity for the octanol phase that their respective (partly) ionized counterparts (Table 1). However, in the case of the acidic ionizable isoprenoids (daidzein, 8-prenylnaringenin, cucurbitacin E) as well as some mycotoxins (β-zearalenol, sterigmatocystin, zearalenone) relatively high partitioning (log $D_{ow} > 1.4$) was observed, although a substantial fraction of the analyte was ionized (> 39%). This illustrates that partitioning of the ionized species cannot always be neglected.

The data generated for those toxins where both single species systems (fully ionized or neutral) were evaluated can be used for modeling of partitioning behavior over the whole pH range for ionizable compounds (Figure 2). Such calculations are based on the fact that a compound’s log $D_{ow}$ is independent of pH as long as only one species exists in the system. A compound’s speciation at any pH can be derived by applying Equation 2 and 3. It is not possible to calculate a compound’s log $D_{ow}$ over the whole pH range when only one species is considered though, as the total difference between log $K_{ow}$ of the neutral and log $D_{ow}$ of the fully ionized species varies from compound to compound. One example are the two cyclic pyrrolizidine alkaloids retrorsine (RET) and senkirkine (SEK) for which an absolute log $D_{ow}$ shift of 2.9 (RET) compared to 1.4 (SEK) is observed when considering the change from fully ionized to neutral form (Figure 2a). For the two open-chain pyrrolizidine alkaloids echimidine (ECH) and lasiocarpine (LAS) the absolute difference is comparable at log $D_{ow} \sim 2.2$ (Figure 2b). Both compounds have highly similar molecular structures, the higher affinity of LAS towards octanol can be explained by the difference of one specific molecular substituent; one hydroxy group (-OH) in ECH is replaced by the less polar methoxy group (-O-CH$_3$) in LAS. According to the observed pH trends for the examples displayed in Figure 2, the experimental data suggest that p$K_a$ values for the cyclic pyrrolizidine alkaloids (Figure 2a) are most likely underpredicted by ACD/Percepta as the inflection point of the fitted curves does not match the p$K_a$ which can be observed for the open-chain pyrrolizidine alkaloids (Figure 2b). Thus, the fitted trends should only be considered for visualization purposes.
Figure 2: Observed log $D_{ow}$ in the range from pH 4.0 to pH 10.2 (grey shaded area) for exemplary basic ionizable compounds (RET, retrorsine; SEK, senkirkine; ECH, echimiidine; LAS, lasiocarpine). For visualization purposes only, trend lines are fitted to experimental data to show the pH dependence of log $D_{ow}$ over the whole pH range from 0 to 14 considering the compound’s ionization as attained by Equation 2. Vertical lines show the predicted pK$_a$ valid for the corresponding cation of the analyzed bases.

**Assessment of Prediction Model Applicability**

Predicted log $K_{ow}$ values for those natural toxins analyzed in their neutral state are displayed in relation to obtained experimental values in Figure 3. Analytes are subdivided into acidic ionizable, basic ionizable and neutral compounds. To evaluate overall model applicability and predictive power for different subgroups, MAE between predicted and experimental data as well as slope and intercept of regression lines are considered (Figure 3, Figure SI-4 to 6). Combining all analytes into one large set for each of the prediction models, data is only evenly distributed around the line of agreement between predicted and experimental data (1:1 line) for the Chemicalize model (Figure 3c). However, slopes of regression lines are close to one for all models (0.92 - 0.99, Figure SI-4 to 6) and intercepts only indicate slight underprediction of partitioning into octanol (intercept -0.26 to -0.1, Figure SI-4 to 6). In KOWWIN and ACD/Percepta about two thirds of all compounds fall within a range of variation of ± 1 log unit.
(KOWWIN 67%, ACD/Percepta 69%), while in case of Chemicalize 89% are well predicted (Figure 3).

Figure 3: Results of the three different QSPR prediction models plotted against experimentally derived log $K_{ow}$ values grouped according to different ionizability of analytes. Only compounds measured in their neutral state are considered (see Table 1). The inner dashed line represents the 1:1 line of agreement between predictions and experiments, while the outer lines indicate an error range of ±1 log unit. MAE is the mean absolute error between all experimental as well as predicted values and n the number of values considered in evaluation.

Chemicalize predicts all subgroups equally well with slopes close to one and the lowest overall MAE (Figure 3c; Figure SI-6). For both other QSPR models, tendencies towards limited applicability for ionizable compounds can be observed. In KOWWIN both acidic (slope = 1.24) and basic (slope = 0.66) ionizable compounds seem poorly predicted, while in ACD/Percepta a similar trend can be seen for basic ionizable compounds (slope = 1.14) only (Figure 3a, b; Figure SI-4 and 5). It has been previously shown, that multifunctional molecules with several ionizable groups are not well predicted by those two models. However, as measurement errors can be rather large (maximum SD = 0.48 for a compound measured in neutral form, Figure SI-3) and some compounds considered for comparison may not be in their neutral state as their predicted pKa is erroneous, this observation should only be seen as a tendency.

In summary, the absolute predictive power of the investigated QSPR models increases from EPISuite's KOWWIN to ACD/Percepta and Chemicalize. Similar observations were made in previous studies comparing different prediction models and experimental data based on larger compound sets. Most methods assume a neutral state of the compounds for calculation,
thus, differences in predicted values may be a result of how charge is generally handled. Additionally, the slightly better performance of ACD/Percepta in comparison with KOWWIN may be due to the larger number of fragments and correction factors considered in ACD/Percepta. All models under investigation are based on atom/fragment contribution methods introducing correction factors to e.g., account for interactions between individual fragments. Only three commonly used, easily accessible QSPR models were considered here while the number of publicly available models is constantly increasing. Thus, to evaluate general performance of individual models, those with other underlying mechanisms should be included. Property based methods such as linear solvation energy relationships (LSER) or methods relying on 3D structures of molecules (e.g., COSMO-RS) may show increased predictive power, particularly for ionizable, multifunctional analytes as those considered in this study.

Conclusions

Overall, experimental data is of great value for describing partitioning behavior of ionizable natural toxins. The miniaturized shake-flask approach as such can be used as a reliable method for the determination of log $D_{ow}$ values for polar compounds. The theoretical operation range of $-1 < \log D_{ow} < 2.5$ could be extended to include natural toxins with a log $D_{ow}$ as low as -1.87 as well. Experimental errors are on average smaller than those introduced by commonly used QSPR models. However, the differences between errors of experiments and predictions are only minor. Thus, QSPR models can be used to narrow down the number of analytes by screening large diverse sets of compounds to prioritize those with properties most relevant for further evaluation. This saves a substantial amount of time and resources in comparison to experimental evaluation of partitioning behavior. However, experiments allow more detailed insights into the partitioning behavior and would particularly be of added value when the focus is on one specific group of compounds only.

For the evaluation of environmental mobility and sorption behavior of polar, ionizable compounds, $K_{ow}/D_{ow}$ has been criticized several times and other intrinsic compound properties
such as the organic carbon-water partition coefficient ($K_{oc}$) were suggested to be a better, more reliable indicator of aquatic mobility.\textsuperscript{46, 47} Firstly, $K_{ow}$ is not always proportional to e.g., $K_{oc}$ for polar molecules, particularly for those that show additional specific (polar) and/or nonspecific (apolar) interactions.\textsuperscript{48} Secondly, as a proxy for sorption behavior it largely neglects ionic interactions of ionizable analytes both with ions in solution and ionizable surface functional groups.\textsuperscript{46} However, $K_{ow}/D_{ow}$ gives reliable indications on a compound’s polarity and thus can be used to further explain observed interactions in sorption studies or as input for bioaccumulation and toxicity models.

\textbf{Acknowledgements}

The authors particularly want to thank Inés Rodríguez Leal (Stockholm University) for providing ACD/Percepta predicted data and input on the evaluation of different models as well as Felix Wettstein (Agroscope) for continuous support with all HPLC related work.

\textbf{Funding}

This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No. 722493 (NaToxAq).

\textbf{Supporting Information}

Details on chemical properties (e.g., CAS number, molecular structure, supplier) and experimental setup (e.g., $\lambda$ of measurement, octanol-water ratios, details on preparation of buffer solutions); literature and predicted log $K_{ow}$ data; results of error analysis; results of regression analysis of each individual QSPR model.
References

(1) Tülp, H.C., Goss, K.U., Schwarzenbach, R.P., and Fenner, K.; Experimental Determination of LSER Parameters for a Set of 76 Diverse Pesticides and Pharmaceuticals. *Environ. Sci. Technol.* 2008, 42, 2034–2040.

(2) Bronner, G. and Goss, K.U.; Predicting sorption of pesticides and other multifunctional organic chemicals to soil organic carbon. *Environ. Sci. Technol.* 2011, 45, 1313-1319.

(3) Bjarnholt, N., Laegdsmand, M., Hansen, H.C., Jacobsen, O.H., and Moller, B.L.; Leaching of cyanogenic glucosides and cyanide from white clover green manure. *Chemosphere* 2008, 72, 897-904.

(4) Clauson-Kaas, F., Ramwell, C., Hansen, H.C., and Strobel, B.W.; Ptaquiloside from bracken in stream water at base flow and during storm events. *Water Res.* 2016, 106, 155-162.

(5) Hoerger, C.C., Wettstein, F.E., Hungerbühler, K., and Bucheli, T.D.; Occurrence and Origin of Estrogenic Isoflavones in Swiss River Waters. *Environ. Sci. Technol.* 2009, 43, 6151–6157.

(6) Kolpin, D.W., Hoerger, C.C., Meyer, M.T., Wettstein, F.E., Hubbard, L.E., and Bucheli, T.D.; Phytoestrogens and Mycotoxins in Iowa Streams: An Examination of Underinvestigated Compounds in Agricultural Basins. *J. Environ. Qual.* 2010, 39, 2089-2099.

(7) Schenzel, J., Hungerbuhler, K., and Bucheli, T.D.; Mycotoxins in the environment: II. Occurrence and origin in Swiss river waters. *Environ. Sci. Technol.* 2012, 46, 13076-13084.

(8) Wink, M.; Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* 2003, 64, 3-19.

(9) Wink, M.; Mode of action and toxicology of plant toxins and poisonous plants, in Wirbeltierforschung in der Kulturlandschaft - Grundlagen und Anwendung. 2009: Berlin, Germany.
(10) Seawright, A.A.; Directly Toxic Effects of Plant Chemicals Which May Occur in Human and Animal Foods. *Nat. Toxins* **1995**, *3*, 227-232.

(11) Ames, B.N., Profet, M., and Gold, L.S.; Nature's chemicals and synthetic chemicals: Comparative toxicology. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 7782-7786.

(12) Ames, B.N., Profet, M., and Gold, L.S.; Dietary pesticides (99.99% all natural). *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 7777-7781.

(13) Teuscher, E. and Lindequist, U.; Biogene Gifte. 3rd Edition ed. **2010**, Stuttgart, Germany: Wissenschaftliche Verlagsgesellschaft. 963 pp.

(14) Schwarzenbach, R.P., Escher, B.I., Fenner, K., Hofstetter, T.B., Johnson, C.A., von Gunten, U., and Wehrli, B.; The challenge of micropollutants in aquatic systems. *Science* **2006**, *313*, 1072-1077.

(15) ECETOC; Environmental Exposure Assessment of Ionizable Organic Compounds - Technical Report No. 123. **2013**, European Centre for Ecotoxicology and Toxicology of Chemicals: Brussels, Belgium.

(16) Schwarzenbach, R.P., Gschwend, P.M., and Imboden, D.M.; Environmental Organic Chemistry. Second Edition ed. **2003**, Hoboken, New Jersey, US: Wiley-Interscience. 1313 pp.

(17) Klöpffer, W.; Verhalten und Abbau von Umweltchemikalien: Physikalisch-chemische Grundlagen. 2 ed. **2012**, Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. 588 pp.

(18) Nolte, T.M. and Ragas, A.M.; A review of quantitative structure-property relationships for the fate of ionizable organic chemicals in water matrices and identification of knowledge gaps. *Environ. Sci. Process Impacts* **2017**, *19*, 221-246.

(19) Rothwell, J.A., Day, A.J., and Morgan, M.R.A.; Experimental Determination of Octanol−Water Partition Coefficients of Quercetin and Related Flavonoids. *J. Agric. Food Chem.* **2005**, *53*, 4355–4360.

(20) OECD; 107: Partition Coefficient (n-octanol/water): Shake Flask Method. *OECD Guidelines for the Testing of Chemicals* **1995**.
OECD; 117: Partition Coefficient (n-octanol/water), High Performance Liquid Chromatography (HPLC) Method. OECD Guidelines for the Testing of Chemicals 2004.

Eadsforth, C.V.; Application of Reverse-phase HPLC for the Determination of Partition Coefficients. Pest. Sci. 1986, 17, 311-325.

Valkó, K., Bevan, C.D., and Reynolds, D.P.; Chromatographic Hydrophobicity Index by Fast-Gradient RP-HPLC: A High-Throughput Alternative to log P/log D. Anal. Chem. 1997, 69, 2022-2029.

Zhao, Y., Jona, J., Chow, D.T., Rong, H., Semin, D., Xia, X., Zanon, R., Spancake, C., and Maliski, E.; High-throughput logP measurement using parallel liquid chromatography/ultraviolet/mass spectrometry and sample-pooling. Rapid Commun. Mass Spectrom. 2002, 16, 1548-1555.

Eadsforth, C., Adams, C., Austin, T., Corry, T., Forbes, S., and Harris, S.; Validation of an HPLC Method for Determining log Pow Values of Surfactants. Tenside Surf. Det. 2014, 51, 230-239.

Poole, C.F., Ahmed, H., Kiridena, W., DeKay, C., and Koziol, W.W.; Contribution of Steric Repulsion to Retention on an Octadecylsiloxane-Bonded Silica Stationary Phase in Reversed-Phase Liquid Chromatography. Chromatographia 2005, 62, 553-561.

Paschke, A., Neitzel, P.L., Walther, W., and Schüürmann, G.; Octanol/Water Partition Coefficient of Selected Herbicides: Determination Using Shake-Flask Method and Reversed-Phase High-Performance Liquid Chromatography. J. Chem. Eng. Data 2004, 49, 1639-1642.

Valkó, K.; Separation methods in drug synthesis and purification. Handbook of analytical separations, ed. Valkó, K. Vol. 1. 2000, Amsterdam, the Netherlands: Elsevier. 589 pp.

Péhourcq, F., Thomas, J., and Jarry, C.; A Microscale HPLC Method for the Evaluation of Octanol-Water Partition Coefficients in a Series of New 2-Amino-2-Oxazolines. J. Liq. Chromatogr. Related Technol. 2000, 23, 443-453.
(30) Andres, A., Roses, M., Rafols, C., Bosch, E., Espinosa, S., Segarra, V., and Huerta, J.M.; Setup and validation of shake-flask procedures for the determination of partition coefficients (logD) from low drug amounts. *Eur. J. Pharm. Sci.* **2015**, 76, 181-191.

(31) Low, Y.W., Blasco, F., and Vachaspati, P.; Optimised method to estimate octanol water distribution coefficient (logD) in a high throughput format. *Eur. J. Pharm. Sci.* **2016**, 92, 110-116.

(32) Bucheli, T.D.; Phytotoxins: environmental micropollutants of concern? *Environ. Sci. Technol.* **2014**, 48, 13027-13033.

(33) Gunthardt, B.F., Hollender, J., Hungerbuhler, K., Scheringer, M., and Bucheli, T.D.; Comprehensive Toxic Plants-Phytotoxins Database and Its Application in Assessing Aquatic Micropollution Potential. *J. Agric. Food. Chem.* **2018**, 66, 7577-7588.

(34) Schenzel, J., Goss, K.U., Schwarzenbach, R.P., Bucheli, T.D., and Droge, S.T.; Experimentally determined soil organic matter-water sorption coefficients for different classes of natural toxins and comparison with estimated numbers. *Environ. Sci. Technol.* **2012**, 46, 6118-6126.

(35) Schwarzenbach, R.P., Egli, T., Hofstetter, T.B., von Gunten, U., and Wehrli, B.; Global Water Pollution and Human Health. *Annu. Rev. Environ. Resour.* **2010**, 35, 109-136.

(36) Arp, H.P.H., Brown, T.N., Berger, U., and Hale, S.E.; Ranking REACH registered neutral, ionizable and ionic organic chemicals based on their aquatic persistency and mobility. *Environ. Sci. Process Impacts* **2017**, 19, 939-955.

(37) Goss, K.U., Arp, H.P., Bronner, G., and Niederer, C.; Nonadditive Effects in the Partitioning Behavior of Various Aliphatic and Aromatic Molecules. *Environ. Toxicol. Chem.* **2009**, 28, 52-60.

(38) US EPA; Estimation Programs Interface Suite™ for Microsoft® Windows, 4.11, available via [https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface.](https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface.) **2017**, United States Environmental Protection Agency, Washington DC, USA.
ACD/Labs; Percepta Platform, release 2016.2, available via https://www.acdlabs.com/products/percepta/. Advanced Chemistry Development, Inc., Toronto, Canada.

ChemAxon; Chemicalize, available via https://chemicalize.com/. ChemAxon Ltd., Budapest, Hungary.

Dohta, Y., Yamashita, T., Horiike, a., Nakamura, a., and Fukami, T.; A System for LogD Screening of 96-Well Plates Using a Water-Plug Aspiration/Injection Method Combined with High-Performance Liquid Chromatography-Mass Spectrometry. Anal. Chem. 2007, 79, 8312-8315.

Westall, J., Leuenberger, C., and Schwarzenbach, R.P.; Influence of pH and ionic strength on aqueous-nonaqueous distribution of chlorinated phenols. Environ. Sci. Technol. 1985, 19, 1360-1367.

Kah, M. and Brown, C.D.; LogD: lipophilicity for ionisable compounds. Chemosphere 2008, 72, 1401-1408.

Machatha, S.G. and Yalkowsky, S.H.; Comparison of the octanol/water partition coefficients calculated by ClogP, ACDlogP and KowWin to experimentally determined values. Int. J. Pharm. 2005, 294, 185-192.

Mannhold, R., Poda, G.I., Ostermann, C., and Tetko, I.V.; Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds. J. Pharm. Sci. 2009, 98, 861-893.

Reemtsma, T., Berger, U., Arp, H.P., Gallhard, H., Knepper, T.P., Neumann, M., Quintana, J.B., and Voogt, P.; Mind the Gap: Persistent and Mobile Organic Compounds-Water Contaminants That Slip Through. Environ. Sci. Technol. 2016, 50, 10308-10315.

Neumann, M. and Schliebner, I.; Protecting the sources of our drinking water - A revised proposal for implementing criteria and an assessment procedure to identify Persistent, Mobile and Toxic (PMT) and very Persistent, very Mobile (vPvM) substances registered under REACH. 2017, German Environment Agency: Dessau-Roßlau, Germany.
(48) Goss, K.U. and Schwarzenbach, R.P.; Linear Free Energy Relationships Used To Evaluate Equilibrium Partitioning of Organic Compounds. *Environ. Sci. Technol.* **2001**, *35*, 1-9.
Experimental Determination of Octanol-Water Partition Coefficients of Selected Natural Toxins

Carina D. Schönsee, Thomas D. Bucheli