UC Riverside
UC Riverside Previously Published Works

Title
Retrieval, Selection, and Evaluation of Chemical Property Data for Assessments of Chemical Emissions, Fate, Hazard, Exposure, and Risks

Permalink
https://escholarship.org/uc/item/762944f5

Journal
ACS Environmental Au, 2(5)

ISSN
2694-2518

Authors
Li, Li
Zhang, Zhizhen
Men, Yujie
et al.

Publication Date
2022-09-21

DOI
10.1021/acsenvironau.2c00010

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

ACS ENVIRONMENTAL Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/environau Review

Retrieval, Selection, and Evaluation of Chemical Property Data for Assessments of Chemical Emissions, Fate, Hazard, Exposure, and Risks

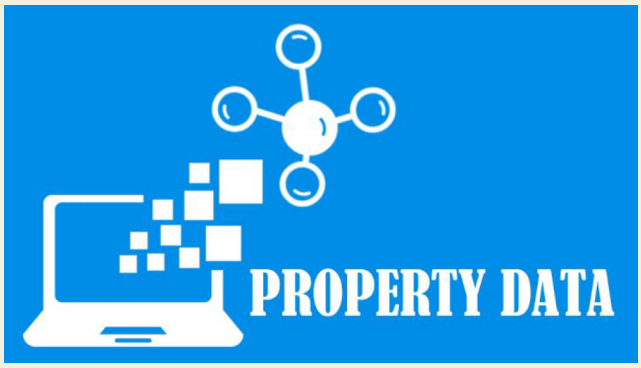
Li Li,* Zhizhen Zhang, Yujie Men, Sivani Baskaran, Alessandro Sangion, Shenghong Wang, Jon A. Arnot, and Frank Wania

Cite This: *ACS Environ. Au* 2022, 2, 376−395

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Reliable chemical property data are the key to defensible and unbiased assessments of chemical emissions, fate, hazard, exposure, and risks. However, the retrieval, evaluation, and use of reliable chemical property data can often be a formidable challenge for chemical assessors and model users. This comprehensive review provides practical guidance for use of chemical property data in chemical assessments. We assemble available sources for obtaining experimentally derived and in silico predicted property data; we also elaborate strategies for evaluating and curating the obtained property data. We demonstrate that both experimentally derived and in silico predicted property data can be subject to considerable uncertainty and variability. Chemical assessors are encouraged to use property data derived through the harmonization of multiple carefully selected experimental data if a sufficient number of reliable laboratory measurements is available or through the consensus consolidation of predictions from multiple in silico tools if the data pool from laboratory measurements is not adequate.



KEYWORDS: *chemical property, partition ratio, half-life, quantitative structure−activity relationship, chemical assessment, modeling, risk, hazard*

INTRODUCTION

Chemical properties are fundamental prerequisites to assessments of chemical emissions, fate, hazard, exposure, and risks. For instance, numeric cut-offs of equilibrium phase partition ratios (also known as partition coefficients) and reaction half-lives are recommended as bright-line criteria in the regulatory assessment of a chemical's potential to resist biotic and abiotic degradation ("persistence"), to accumulate in organisms ("bioaccumulation"), to permeate through natural barriers such as soils, riverbanks, and aquifers ("mobility"), and for global transport to remote areas ("long-range environmental transport").[1−6] Combining these separate criteria can inform the assessment of the overall inherent hazard of chemicals and prioritize top-ranked chemicals for more stringent scrutiny and regulation. Also, chemical properties constitute critical input information to chemical assessment models, which evaluate chemical emissions, fate, and transport in indoor and natural environments, as well as the exposure to ecological receptors and human beings.[7−9]

Reliable chemical property data are the key to defensible and unbiased chemical assessments because chemical behaviors in the environment and organisms can be sensitive to minor differences in their properties. A well-known example is the distinct mode of transport of hexachlorocyclohexane (HCH) isomers from source regions to Northern Oceans: α-HCH is largely atmospherically transported, but β-HCH is transported mainly through oceanic currents, although α-HCH's air−water partition ratio (K_{AW}) is merely an order of magnitude higher than that of β-HCH.[10] Chemical properties such as the octanol−water (K_{OW}) and octanol−air (K_{OA}) partition ratios also inform hazard assessment and classifications derived from strict "cut-off" criteria. For instance, a persistent chemical is considered to have bioaccumulation potential in water-respiring organisms if it has a K_{OW}, e.g., greater than 10^5 according to the Stockholm Convention or $10^{4.5}$ according to the European Chemicals Agency (ECHA),[2,11] or in air-respiring organisms if it has a K_{OW} greater than 10^2 and a

Received: March 2, 2022
Revised: July 1, 2022
Accepted: July 5, 2022
Published: July 19, 2022

ACS Publications
© 2022 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

K_{OA} greater than 10^5.[12,13] In some cases, even property values with uncertainties of many orders of magnitude may be sufficient, for example, if such an uncertain value is far from a regulatory threshold value of interest or if an investigated fate process is insensitive to the property within the range of its uncertainty. In model-based assessments, the uncertainty associated with chemical input properties can be propagated to model predictions and, by extension, the decision-making process. Zhang et al.[14] show that a given chemical's assessed potential for bioaccumulation and long-range transport can differ depending on which property prediction tools are used. Similarly, Odziomek et al.[15] found that a model yields widely divergent values for persistence and long-range environmental transport potential if the input K_{AW} is calculated from estimated K_{OA} and K_{OW} or if it is calculated from an estimated Henry's law constant (see section Property Data in Chemical Assessments for the two ways of calculations).

Even though the decisive role of chemical properties in chemical assessments has long been recognized and widely acknowledged, the retrieval, evaluation, and use of reliable chemical property data remain a formidable challenge. First, although experimentally derived data are generally given the highest priority in regulatory assessment practices,[16−18] measurements performed in different laboratories or based on different techniques are often associated with different levels of uncertainty, reliability, and consistency. One example is the remarkable variability in measured water solubility and K_{OW} of dichlorodiphenyltrichloroethane (DDT) reported in the literature, spanning up to 4 orders of magnitude due to the abundance of biased and even erroneous measurements or flawed reporting.[19] Variabilities of over 3 orders of magnitude are also evident in chemicals documented in the International Uniform Chemical Information Database (IUCLID).[20] Clearly, experimentally derived data must be critically analyzed, reviewed, and harmonized before they are suited for chemical assessment purposes.[9,19] While critical evaluation of properties has been recognized and emphasized for decades, current chemical assessment practices often continue to ignore the reliability of experimentally derived data, since regulatory efforts lack formal property evaluation guidance and build on a seemingly random or arbitrary selection from the available experimental data without an in-depth evaluation of the associated uncertainties.

Furthermore, experimentally derived data are incomplete or missing for most chemicals used in commerce, a situation that is even more pronounced for premanufactured chemicals or chemicals in design. The assessment of these "data-poor" chemicals necessitates the use of various in silico prediction tools, such as empirical correlations including single- and poly parameter linear free energy relationships (sp- or pp-LFERs), and quantitative structure−activity or −property relationships (QSARs or QSPRs), whose reliability and predictability are generally statistically established through a series of internal and external validations.[21] Approaches based on quantum chemistry and statistical thermodynamics, which calculate chemical properties by considering solute−solution interactions and require no specific calibration, have also been developed.[22] However, some regulatory and academic communities have yet to fully embrace the use of in silico predictions, possibly due to a widely held belief that despite their potential flaws and errors, experimentally obtained data are more trustworthy and always superior to in silico predictions. No single QSAR or QSPR can be the best at predicting chemical properties for all chemicals because different tools build on distinct training sets, employ different prediction algorithms, and therefore demonstrate suitability for different parts of the chemical space.[23] Reliable and suitable predictions from different in silico tools can be combined and harmonized before their use in chemical assessments. Preliminary attempts have been made by several integrated online platforms, such as CompTox (www.comptox.epa.gov) developed by the U.S. Environmental Protection Agency and the Exposure And Safety Estimation (EAS-E) Suite (www.eas-e-suite.com) developed by ARC Arnot Research & Consulting.

The current literature lacks a critical review guiding the comprehensive, systematic evaluation of the availability, quality, and consistency of experimentally derived and in silico predicted property data for chemical assessments. For many academic or regulatory chemical assessors, answers to the following three questions are still missing pieces of the puzzle:

(1) From where and how can we retrieve experimentally derived or computationally predicted chemical property data suited for chemical assessments?

(2) How should we evaluate the quality and reliability of the retrieved data?

(3) How can we aggregate and harmonize experimentally derived or computationally predicted chemical property data retrieved from different sources?

To address these issues, we here seek to (i) assemble sources for retrieving chemical properties for chemical assessments, (ii) provide general principles for evaluating the quality and reliability of the retrieved property data, and (iii) demonstrate how the quality and reliability of property data can be improved by the harmonization of experimentally derived data from multiple sources or in silico predictions from multiple prediction tools.

■ PROPERTY DATA IN CHEMICAL ASSESSMENTS

We limit the scope of this review to pure organic chemicals with discrete structures. We focus on two categories of chemical properties that are required by most chemical assessments, namely partition ratios and rates of reactions. We begin with a short overview of the definitions and applications of these properties. For more information about these properties, we refer readers to Boethling et al.[24]

Partition ratios (K_{ij}) describe the equilibrium distribution of chemicals between two immiscible phases i and j.[25] Previously, these equilibrium distribution ratios have been commonly referred to as partition coefficients.[26] However the International Union of Pure and Applied Chemistry (IUPAC) has recommended the use of "partition ratios" because these properties are dependent on both the chemical, the solvent phases, and the temperature of the system.[27]

These partition ratios can inform the phase distribution under environmentally and biologically relevant conditions.[26] For instance, partition ratios between neutral storage lipids and water ($K_{nlip\text{-}water}$), between phospholipids and water ($K_{plip\text{-}water}$), and between proteins and water ($K_{protein\text{-}water}$) characterize the distribution and accumulation of chemicals within organisms; organic carbon−water (K_{OC}) partition ratios quantify the sorption of chemicals by natural organic matter from water in the environment. Underlying the use of partition ratios is an assumption of linear phase distribution or sorption, which is often the case at low concentrations.

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

Notably, by using water as a proxy for polar aqueous phases (e.g., water and mucus) and *n*-octanol as a proxy for weakly bipolar organic condensed phases (e.g., lipids, lipid-like plant constituents, as well as the organic fractions of dust, soils, sediments, and polymeric materials), one can approximate various phase distribution processes using three volume-based, unitless equilibrium partition ratios in a well-defined air (A)−water (W)−octanol (O) system, namely, K_{OA}, K_{OW}, and K_{AW}. These partition ratios, also known as "partition constants",[26] can be measured in the laboratory with standard, reproducible methods. The "three solubilities approach"[28,29] expresses the three partition ratios as ratios of solubilities in air (S_A, which is the saturation vapor pressure V_P divided by the product of gas constant R and temperature T), water (that is, the aqueous solubility or S_W), and octanol (S_O). For example, K_{OW} is defined as the ratio of S_O to S_W, and K_{AW} is defined as the ratio of S_A to S_W. This entails that the three partition ratios are related via a thermodynamic triangular relationship,[28,29] which allows any one of the three coefficients to be estimated from the other two, that is, $K_{OW} = K_{OA} \times K_{AW}$.

In many instances, air−water partitioning is also described by the Henry's law constant or Henry's law coefficient in water (H_V^{pc}) with units of Pa m^3 mol^{-1}, which is K_{AW} multiplied with the product of R and T. Solvent-air partitioning can also be described by solvent-specific Henry's law constants or coefficients; for instance, octanol−air partitioning is related to the Henry's law coefficient in octanol.[30] Readers are directed to Sander et al.[31] for the specifics of the naming convention and use of Henry's law constant.

Partition ratios and solubilities are functions of temperature. One can convert their values at a given temperature (e.g., 25 °C measured in the laboratory) to another (e.g., an environmentally relevant temperature[32]) using an internal energy ($\Delta U°$; when the abundance in the gas phase at a given temperature is defined on a volumetric concentration basis) or enthalpy ($\Delta H°$; when this abudance is defined on a partial pressure basis) of phase change and a version of the van't Hoff equation (the Clausius−Clapeyron equation or the empirical Antoine equation).[26,33]

For ionizable chemicals, partition ratios involving the aqueous phase also depend on the extent of dissociation, which is a function of a chemical's acid or base dissociation constant (pK_a or pK_b) and the environmental pH according to the Henderson−Hasselbalch equation.[26] A small pK_a or pK_b represents a strong acid or base, and a larger difference between the environmental pH and pK_a or pK_b indicates a greater extent of dissociation. Dissociation favors partitioning into the water phase and disfavors partitioning into the gas phase because ions are more water-soluble and less volatile than their neutral counterpart. Therefore, the partitioning behavior of these chemicals is often described by an abundance-weighted average of the partition ratios of their neutral and charged forms, referred to specifically as distribution ratios (D_{ij}).[26,34]

Rate constants (k) are used to quantify the rates of chemical reaction or removal from the environment or organisms, through processes such as oxidation by the hydroxyl radical in indoor and outdoor air, microbial biodegradation, and metabolic biotransformation in higher-level organisms. First-order rate constants (with units of reciprocal time) are most common in chemical assessments. Take biochemical reactions as an example: When chemical concentrations are lower than the Michaelis−Menten constant K_M (a measure of enzyme−substrate binding affinity), which is typical at low contaminant exposures, the Michaelis−Menten kinetics can be adequately approximated by the first-order reaction. Furthermore, the delivery of the environmental contaminant to the site of enzymatic activity is often the rate-limiting step, which then follows first-order kinetics. Second-order rate constants are also used for certain reactions: for instance, the rate constant for reaction with the hydroxyl radical (in cm^3 molecule^{-1} s^{-1}) is often multiplied with the concentration of the hydroxyl radical, and hydrolysis rate constants (in L mol^{-1} s^{-1}) are often combined with the concentration of [H$^+$] or [OH$^-$] in water (i.e., the water's pH).

Chemical assessments also utilize the concepts "reaction half-life" and "reactive residence time", which both have a unit of time and are often more intuitive. Half-lives describe the time required for removing one-half of a given amount of chemical and are quantified as $\ln 2/k$ for first-order reactions. Residence times describe the average time that a given amount of chemical resists a reaction and is quantified as the reciprocal of k. For instance, the biodegradation half-life (HL_{biodeg}) and biotransformation half-life ($HL_{B,fish}$ for fish and $HL_{B,human}$ for humans) are input parameters to most fate, exposure, and bioaccumulation models. Exposure science and toxicokinetics also use the concept of "clearance" to describe the rate of chemical reaction or removal; clearance is defined as the equivalent volume of blood or plasma that can be cleared of a chemical within a given unit of time. Rate constants can also be dependent on temperature. One can convert the rate constant at a given temperature to another using the activation energy of the reaction and the Arrhenius equation.[26] A high Arrhenius activation energy indicates a greater sensitivity of the rate constant to temperature.

SOURCES FOR RETRIEVAL OF CHEMICAL PROPERTIES

Laboratory Measurements

Laboratory measurements are the most fundamental approach for obtaining chemical properties. To date, experimentally derived data for thousands of chemicals, mostly well-investigated pharmaceuticals and environmental contaminants such as organochlorines and organophosphates, have been compiled and documented in comprehensive reviews (such as refs. 30, 35, 36; not exhaustive), handbooks (such as refs 34, 37−41; not exhaustive) and online databases (such as REACH registered substances database, the LOGKOW database by Technical Database Services, the Henry's law constant database,[42] the PhysProp database by Syracuse Research Corporation, and those built in the integrated online platforms as elaborated later; not exhaustive) for assessment use.

While experimentally derived data are often recommended to be given the highest priority in chemical assessment practice,[16−18] this must be done with caution because not all laboratory measurements are of the same quality, and therefore, they are not equally credible or reliable.[24,43] Typical indications for low data quality include but are not limited to

(1) the use of nonstandard methodologies or noncompliance with Good Laboratory Practices
(2) inadequate documentation of the measurement procedure, test chemical, or biological material[19]
(3) inconsistency between the chemicals used in the test and reported in the test results (such as pure chemical vs technical substance or isomeric mixture, neutral acid/

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

base vs electronically charged salt, parent chemical vs metabolites, liquid vs solid states at the temperature of interest)[19,24]

(4) an insufficient number of independent experimental runs or a lack of replication

(5) the measured properties being outside of the range of applicability or suitability of an experimental design/technique or the instrumental analysis method[30]

(6) errors in processing and reporting data, units, and references[19,44]

(7) a lack of appropriate documentation of uncertainty information, such as the absence of the standard deviation of measurements, and statistically insignificant differences between experimental and control groups. A value without information on its uncertainty can even be inferior to a value with high reported uncertainty.

Decades ago, Kollig[43] proposed a series of criteria to gauge the quality of experimental data based on analytical, experimental, statistical, and corroborative aspects.

In general, extreme property values, such as minimal volatility ($V_P < 10^{-5}$ Pa or $K_{OA} > 10^{12}$) or high hydrophobicity ($K_{OW} > 10^5$), are more likely to be uncertain because of the increased complexities of reliably quantifying very small quantities of a chemical in the gas or aqueous phase, respectively. In other cases, kinetic limitations may prevent less volatile or highly hydrophobic chemicals from reaching phase equilibrium during the determination of an equilibrium partition ratio.[30,45] Likewise, biodegradation rates for highly recalcitrant chemicals (half-life > 1 year) can be difficult to determine during short incubation periods because of minuscule differences between the removal rates in experimental and biologically inactivated control groups. When the removal rate is smaller than the uncertainty of the analytical method, such as the typical level of 20−30% in the broadly used liquid chromatography−mass spectrometry/mass spectrometry (LC-MS/MS) techniques in the absence of a suitable internal standard, the estimated biodegradation rate constant or half-life becomes unreliable and should not be used for chemical assessments.[46] A prolonged incubation time and more experimental data points are thus warranted to minimize the interference of analytical variation.

Some laboratory measurements may also be of low quality due to the selection of an inappropriate experimental design/technique or the instrumental analysis method. As a rule of thumb for quick reference, Figure 1 displays the *general* ranges of applicability for commonly used techniques for measuring partition ratios, water solubility, and vapor pressure. It may be possible to expand the displayed ranges of applicability if special considerations or care are taken. Attention needs to be paid during the selection and evaluation of properties reported for chemicals located outside these ranges of applicability.

One may also need to pay special attention to data derived from "indirect" measurements, such as those obtained through correlations with chromatographic retention times (e.g., K_{OA} measurements with gas chromatography or K_{OW} measurements with liquid chromatography) or numbers calculated from measurements of other properties (e.g., log K_{AW} as the difference between independent measurements of log K_{OA} and log K_{OW}).[47] For instance, chromatographic retention time techniques for K_{OA} or K_{OW} often rely on "reference" or "calibration" compounds with reliably known partition ratio values. In such a case, chemical assessors must examine

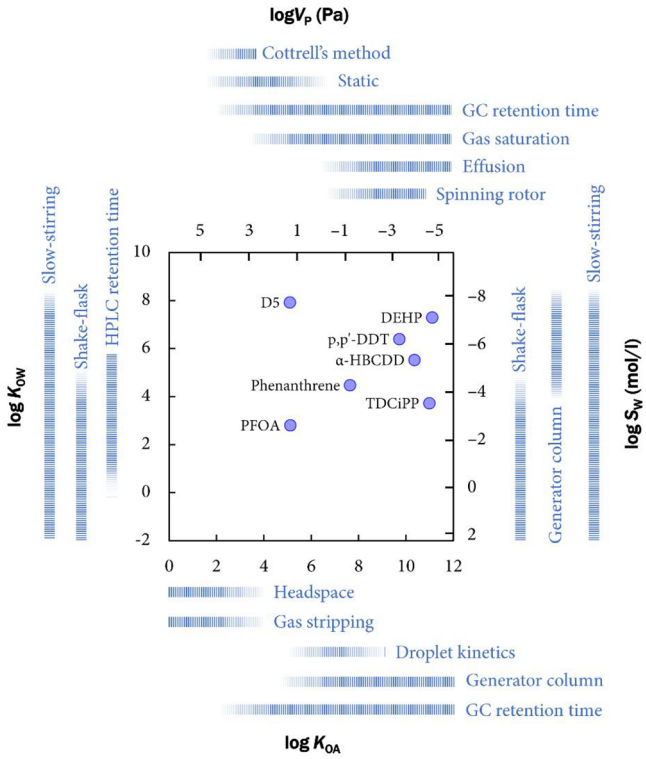


Figure 1. Recommended ranges of applicability of commonly used methods for measuring octanol−water (K_{OW})[60−62] and octanol−air (K_{OA}) partition ratios,[30,63,64] water solubility,[65] and vapor pressure.[66] The scales for water solubility and vapor pressure are calculated from K_{OW} and K_{OA}, respectively, using the empirical equations outlined in Table 2. Also displayed are archetypical chemical contaminants representative of different combinations of partitioning properties: D5 = decamethylcyclopentasiloxane; DEHP = di(2-ethylhexyl) phthalate; DDT = dichlorodiphenyltrichloroethane; HBCDD = hexabromocyclododecane; TDCiPP = tris(1,3-dichloro-2-propyl)phosphate; PFOA = perfluorooctanoic acid (the neutral form is considered here). Their positions are based on the final adjusted values of their K_{OW} and K_{OA}, as reported in Figure 2.

whether the polarity of the reference chemicals is similar to that of the test chemicals and whether the chemical's ability to undergo hydrogen bonding with octanol and the stationary phase is similar.[30] In practice, a chemical can be assessed to be polar if a linear combination of Abraham solute descriptors for hydrogen bonding acidity (A) and basicity (B), weighted by system constants *a* and *b* from corresponding pp-LFER equations (see section Empirical Correlations below), is high.[30,48,49] The check for the similarity in polarity is particularly important if only a single reference compound is used. When multiple reference compounds are involved, the test compound should be located within the range of the partition ratios of reference compounds such that interpolation is used. However, extrapolation is acceptable and even "necessary" for the determination of extremely low or high partition ratios.[50]

Flawed experimental data may be the result of unaccounted sorption processes. For example, Shunthirasingham et al.[51] have demonstrated that the bubble stripping method for K_{AW} determination yields biased results for substances adsorbing to the air−water surface of the bubbles. This process can affect not only surfactants but also sparingly soluble solutes.[52] The resulting artifact can be reduced by reducing the surface-to-

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

volume ratio of the stripping gas through the use of larger bubbles.[53]

Another potential source of error is the loss of chemicals in biodegradation tests due to processes other than biodegradation, such as dilution, biomass adsorption, and evaporation.[54−58] These loss processes can lead to an overestimation of the biodegradation rate constant. Before using data for chemical assessments, one may need to first examine, for example, whether the test was conducted in a closed system, whether a stringent mass balance of the parent compound was obtained,[54,59] whether appropriate techniques were employed to quantify biological utilization and integration of the parent compound into biomass, and whether plausible transformation products were identified.

Table 1 illustrates the considerations to be applied in the selection and curation of experimentally derived data for one example chemical, namely di(2-ethylhexyl)phthalate (DEHP), a plasticizer commonly found in consumer products and building materials. While a wide array of laboratory measurements is available for individual properties, most of them may not be well-suited for a chemical assessment because of failure to meet one or more quality assurance criteria. It is also interesting to note that in Table 1, the majority of excluded values (i.e., marked as "not used") predate the 1990s and were produced before international test standards and guidelines were published. This highlights the fact that "old" experimental data need to be carefully evaluated to verify that the methods and analytical techniques used at the time meet current quality standards. Nevertheless, an evaluation of data quality as strict as that presented in Table 1 is not always possible because, for most chemicals used in commerce, experimentally derived data are scarce. In this case, one may need to compromise by applying looser quality criteria that reflect shortcomings of the available data.

When multiple measurements are available for the same chemical or a suite of structurally similar chemicals (e.g., congeners or homologues), one may also check the consistency between them to identify outliers. For instance, one can examine whether measurements of vapor pressure at different temperatures satisfy the Antoine equation by plotting logarithmic vapor pressure against reciprocal absolute temperature. Data points falling far from a straight line are less likely to be reliable.[37] Also, the properties of individual homologues often exhibit consistent patterns; comparing experimentally derived data between homologues may thus allow for the identification of potentially less reliable measurements.[92,93]

Carefully evaluated and curated data from a single experimental study can be suited for chemical assessments, particularly when the measurements are not too challenging (e.g., do not involve the determination of K_{OW} or S_W for superhydrophobic compounds or K_{OA} for low volatility compounds) and were made in compliance with Good Laboratory Practices using standard methodologies. When multiple experimentally derived data are available, they can be consolidated in literature-derived values (LDVs), with their central tendency described by corresponding means (e.g., geometric or arithmetic).[44]

For properties related to the air−water−octanol system, one needs to additionally check whether their LDVs are thermodynamically consistent. While the "three solubilities approach"[28] defines the relationship between partition ratio and solubilities in octanol, air, and water, such thermodynamic constraints are often violated in reality if the properties are sourced from different independent measurements. It is important to realize that some deviations from thermodynamic constraints are not due to measurement uncertainty but reflect an inherent inconsistency between the experimental systems represented by different measurements. For example, the partial miscibility between water and octanol implies that a measured K_{OW} expresses the equilibrium between water-saturated ("wet") octanol and octanol-saturated water, whereas the K_{AW} and K_{OA} involve equilibria with pure solvents (pure water and "dry" octanol).[34,94] Therefore, if using "wet" octanol for K_{OA} or "dry" octanol for K_{OW} in chemical assessments, one would expect a biased apparent solubility of polar chemicals in organic condensed phases such as lipids, dust, soils, sediments, and polymeric materials. For these reasons, an adjustment to individual partition ratios and solubilities is warranted to ensure the adherence to those fundamental thermodynamic relationships.[9,94] By implementing empirical correlations, such an adjustment converts LDVs to harmonized, thermodynamically consistent values named final adjusted values (FAVs) that diverge minimally from the LDVs.[9,94] The extent of adjustment can be scaled to the estimated relative uncertainty of the LDVs.[9] A useful tool for such an adjustment is the least-squares adjustment procedure by Schenker et al.;[9] it is available as a Microsoft Excel spreadsheet with a friendly graphical user interface.

As an illustrative example, one can calculate the LDVs of V_P (2.16×10^{-5} Pa), S_W (0.0227 mg/L), and K_{OW} ($10^{7.43}$) of DEHP, as the geometric means of laboratory measurements in Table 1 that are evaluated to be reliable and marked with "used". One can then obtain corresponding FAVs (V_P of 2.16×10^{-5} Pa, S_W of 0.0227 mg/L, S_O of 1.56×10^3 mol/m^3, K_{OW} of $10^{7.43}$, K_{OA} of $10^{11.25}$, and K_{AW} of $10^{-3.82}$) using the least-squares adjustment procedure by Schenker et al.[9] In this particular case, FAVs are equal to the LDVs because LDVs for only three properties could be obtained. A comparison between these FAVs of DEHP and the collected measurements in Table 1 indicates that certain laboratory measurements deviate substantially from FAVs, which underscores the importance of independent evaluation of experimentally derived data before applications in chemical assessments.

Following the same approach, we obtain the FAVs of 76 neutral organic chemicals with experimentally derived data sets that are sufficiently large for the derivation of FAVs; these chemicals include 11 organobromines, 34 organochlorines, 3 organophosphates, 4 phthalates, 6 perfluoroalkyl substances (PFASs), 15 polycyclic aromatic hydrocarbons (PAHs), and 3 organosilicons (Figure 2). The LDVs are taken from the literature[92,95−100] (for 60 chemicals) or calculated based on compiled and curated experimentally derived data reported in the literature (for 16 chemicals) (see Supporting Information Tables S1 and S2 for details). The FAVs of the 60 chemicals differ from earlier reported ones because we use a slightly different adjustment procedure.

Note that the derivation of FAVs is often an iterative, rather than a sequential, process. One may first apply the least-squares adjustment procedure to a preliminary set of LDVs and check the direction and magnitude of the adjustment of each LDV. Such information helps identify the LDVs that are remarkably inconsistent with others and, hence, more likely to be questionable. One can then return to the data used in the derivation of the LDVs and check whether the questionable LDVs are biased by inappropriate data point(s) and whether recalculation of these LDVs is needed.

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

Table 1. Illustrative Example of the Selection and Evaluation of Experimentally Derived Data for Di(2-ethylhexyl)phthalate (DEHP) for Chemical Assessments

property	value	reference	remark
log K_{OW}	7.94	Howard et al. (1985)[67]	Not used. Measured using the reversed-phase liquid chronography method with seven nonpolar and lowly polar reference compounds (the Abraham solute descriptor *B* lower than 0.3) with log K_{OW} ranging from 2.13 to 6.72.
	7.86	Harnisch et al. (1983)[68]	Not used. Measured using the reversed-phase liquid chronography method with lowly polar reference compounds of alkylbenzenes (the Abraham solute descriptor *B* lower than 0.3) with log K_{OW} ranging from 3.17 to 7.33.
	5.03	Harnisch et al. (1983)[68]	Not used. Measured as a part of the OECD laboratory comparison tests using the shake-flask method, not suitable for hydrophobic chemicals with log K_{OW} greater than 5.
	7.86 ± 1.33	Klein et al. (1988)[50]	Used. Measured in multiple tests using the reversed-phase liquid chronography method with over 50 polar and nonpolar reference compounds with log K_{OW} ranging from 0.3 to 6.19.
	7.453 ± 0.061	Brooke et al. (1990)[69]	Used. Measured independently by two laboratories using the slow-stirring method (stirred for more than 36 h), suitable for highly hydrophobic chemicals with log K_{OW} up to 8.2.
	7.137 ± 0.153		
	7.453 ± 0.061	De Bruijn et al. (1989)[70]	The same value reported by Brooke et al. (1990). Measured using the slow-stirring method (stirred for more than 36 h), suitable for highly hydrophobic chemicals with log K_{OW} up to 8.2.
	7.27 ± 0.04	Ellington and Floyd (1996)[71]	Used. Measured using the slow-stirring method (stirred for 24 h), suitable for highly hydrophobic chemicals with log K_{OW} up to 8.2.
log K_{OA}	10.48 ± 0.69	Pegoraro et al. (2015)[72]	Not used. Recalibrated based on a previously model-predicted value.
V_P (Pa)	1.9×10^{-5}	Hinckley et al. (1990)[73]	Not used. Measured using the gas chromatography (GC) retention time method based on the nonpolar reference compound C20 alkane, not suitable for polar compounds
	1.9×10^{-5}	Small et al. (1948)[74]	Not used. Extrapolated to 25 °C using an empirical equation, $\log(V_P/\text{mmHg}) = -5757/(T/K)+12.47$, fitted based on vapor pressures at different temperatures much higher than 25 °C (around 100 to 150 °C) determined using the vapor pressure balance effusion method.
	3.0×10^{-5}	Perry and Weber (1949)[75]	Not used. Extrapolated to 25 °C using an empirical equation $\log(V_P/\mu\text{mHg}) = -5440/(T/K)+14.62$, fitted based on vapor pressures at different temperatures (99 to 148 °C) determined using a modified dynamic method (on a pendulum tensimeter).
	8.2×10^{-6}	Werner (1952)[76]	Not used. Extrapolated to 25 °C using an empirical equation $\log(V_P/\mu\text{mHg}) = -3811/((T/°C)+201.2)+12.639$, fitted based on vapor pressures at different temperatures much higher than 25 °C (120 to 225 °C) determined by a modified dynamic method (on a tensimeter).
	8.6×10^{-4}	Howard et al. (1985)[67]	Not used. Measured using the gas saturation method at 25 °C. Measured vapor pressures differ by a factor of 5 between the three gas flow rates used in the experiment. Such dependence on the flow rate indicates suspected unsaturation of the gas, according to the OECD Test Guideline 104.
	1.3×10^{-5}	Gobble et al. (2014)[77]	Used. Measured using the GC retention time method with previously accurately determined vapor pressures of other dialkyl phthalates as reference compounds, in the combination of self-consistency between their vapor pressures and vaporization enthalpies.
	3.6×10^{-5}	Liang and Xu (2014)[78]	Used. Measured using the gas saturation method at 25 °C. While the authors used only one flow rate, instead of three as recommended by the OECD Test Guideline 104, they ensured the saturation of the gas by periodically collecting and analyzing air samples from the outlet of the chamber.
S_W (mg/L)	0.285	Hollifield (1979)[79]	Not used. Measured for technical DEHP product, not pure chemical.
	0.041	Leyder and Boulanger (1983)[80]	Not used. Measured (20 °C) using the shake-flask method followed by centrifugation and gas chromatography quantification. The vigorous agitation used in the shake-flask method can cause emulsion of DEHP because DEHP is liquid, hydrophobic, and viscous. Centrifugation cannot separate DEHP emulsions from dissolved DEHP, because DEHP has a specific gravity close to that of water.[65]
	0.40 ± 0.05	Wolfe et al. (1980)[81]	Not used. Measured using a modified shake-flask method, followed by centrifugation and gas−liquid chromatography quantification. The vigorous agitation used in the shake-flask method can cause emulsion of DEHP because DEHP is liquid, hydrophobic, and viscous. Centrifugation cannot separate DEHP emulsions from dissolved DEHP, because DEHP has a specific gravity close to that of water.[65]
	0.34 ± 0.04	Howard et al. (1985)[67]	Not used. Measured using a modified shake-flask method, followed by centrifugation and high-performance liquid chromatography quantification. The vigorous agitation used in the shake-flask method can cause emulsion of DEHP because DEHP is liquid, hydrophobic, and viscous. Centrifugation cannot separate DEHP emulsions from dissolved DEHP, because DEHP has a specific gravity close to that of water.[65]
	0.36	DeFoe et al. (1990)[82]	Used. Measured using the shake-flask method, followed by indirect determination of turbidity inflection. While the shake-flask method is less suitable for lowly soluble chemicals (<10 mg/L) because of emulsion, such an indirect approach minimizes the impact of DEHP emulsion.
	0.017	Thomsen et al. (2001)[83]	Used. Measured using the shake-flask method, followed by indirect determination of the surface tension. While the shake-flask method is less suitable for lowly soluble chemicals (<10 mg/L) because of emulsion, such an indirect approach minimizes the impact of DEHP emulsion.
	0.0019	Letinski et al. (2002)[84]	Used. Measured using the slow-stirring method, followed by solid-phase extraction and GC quantification. The slow-stir flask method does not cause significant DEHP emulsion. Mercuric chloride was added to inhibit microbial degradation during slow stirring.

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

Table 1. continued

property	value	reference	remark
$t_{1/2,\ mix}$ (day)	4.94 – 5.25	Sugatt et al. (1984)[85]	Not used. No sterile/heat-inactivated controls were included. Measured using the shake-flask method for acclimated culture (from combined soil and activated sludge) based on the release of CO_2. Without isotopic labeling of the carbon, CO_2 measurement can be easily interfered with and might not accurately reflect DEHP biodegradation.
$t_{1/2,\ soil}$ (day)	40	Cartwright et al. (2000)[86]	Not used. Since less than 30% of DEHP was removed over 70 days, which was very close to the removal rate of the sterile control and the detection variation of the analytical method, it is difficult to determine whether such a low removal rate was due to biodegradation or artifacts.
	2 – 5	Juneson et al. (2001)[87]	Not used. Measured based on the comparison with a vaguely described "abiotic control", where the removal of DEHP was claimed to be "negligible" without information on sterilization. Less than 30% of DEHP was removed in some biodegradation tests; the removal rate is close to the detection variation of the GC analytical method. Caution must be exercised if this measurement is used.
	26.3 ± 0.7 30.8 ± 0.7	Xu et al. (2008)[88]	Used. Measured with sterile controls, detailed information on the extraction recovery and the detection limits of DEHP, and the full data set and steps for the half-life calculation.
$t_{1/2,\ sediment}$ (day)	14	Yuwatini et al. (2006)[89]	Not used. No sterile controls were used. Detection limits were provided.
	14.8 (7.3–27.5)	Yuan et al. (2002)[90]	Used. Measured with sterile controls, and detailed information on the extraction recovery and the detection limits of DEHP (the experimental setup was reported to be identical to Yuan et al.[91])
$t_{1/2,\ water}$ (day)	0.42	Yuwatini et al. (2006)[89]	Used. Measured with sterile controls treated with ultraviolet irradiation, where biological activities were largely inhibited. Detection limits were provided.

Recommended Empirical Values

When experimentally derived data are not available for a specific chemical of interest, one may be able to use recommended generic defaults, which reflect the consensus among the scientific community after critically reviewing laboratory measurements for many other chemicals. Such a practice is particularly useful for screening purposes and for properties that are difficult to measure and/or less influential on model outputs. An example is the selection of Arrhenius activation energies. The ECHA recommends an empirical relationship for correcting the hydrolysis half-life for temperature, which corresponds to a generic activation energy of 54 kJ/mol.[101] The European Food Safety Authority recommends a generic, median activation energy of 65.4 kJ/mol to be used for chemical biodegradation in soil.[102] A generic activation energy of 8 to 10 kJ/mol, based on experimental data of PCB congeners,[103] is often assumed for the temperature dependence of gas-phase reactions with the hydroxyl radical.

Generic empirical ratios have also been used to extrapolate degradation half-lives between different environmental media when chemical-specific experimentally derived data are not available. Fenner et al.[104] recommended the use of a ratio of 1:2:10 for the extrapolation of degradation half-lives between water, surface soil (aerobic), and sediment (anaerobic) in screening-level assessments. Also, other extrapolation factors such as 1:1:4 (Boethling et al.[105]) and 1:2:9 (Aronson et al.[106]) have previously been proposed for screening purposes.

It should be cautioned that the use of such generic values is a measure of practical expediency because they merely reflect the central tendency of values of a wide range of chemicals. The "true" values for a specific chemical may deviate substantially from the central tendency. For instance, while internal energies of solvation are often assumed to be 20 and 0 kJ/mol for water (ΔU_W) and octanol (ΔU_O), highly hydrophobic and bulky molecules, e.g., heptachlorodibenzo-*p*-dioxin, may have negative values of ΔU_W.[94] Therefore, the use of generic empirical values is justified only if the data availability is limited and if the range of the value is known to be narrow. And it does not discourage the search for further chemical-specific scientific understanding.

Empirical Correlations

Chemical properties that are environmentally or biologically relevant but often demanding or uncertain to determine, can also be calculated using empirical correlations, based on properties called descriptors that can be more accurately and reliably measured with standard methodologies in laboratories. For instance, a sp-LFER correlates an environmentally or biologically relevant partition ratio with K_{OW}, K_{AW}, or K_{OA}. Underlying the sp-LFER practice is a reductionistic assumption that the partitioning of chemicals between environmental media can be simplified mainly as the partitioning within the air−water-octanol system. Likewise, Fenner et al.[107] proposed an empirical relationship correlating the biodegradation half-life in soil with the biodegradation half-life in active sludge from municipal wastewater treatment plants, given that the latter can be readily determined with standard methodologies. One can also estimate the molar enthalpy or internal energy of vaporization of liquids and subcooled liquids from V_P or K_{OA} of non-hydrogen bonding liquids and subcooled liquids, based on Trouton's rule.[108,109] Table 2 compiles published sp-LFERs for properties widely used in chemical assessments.

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

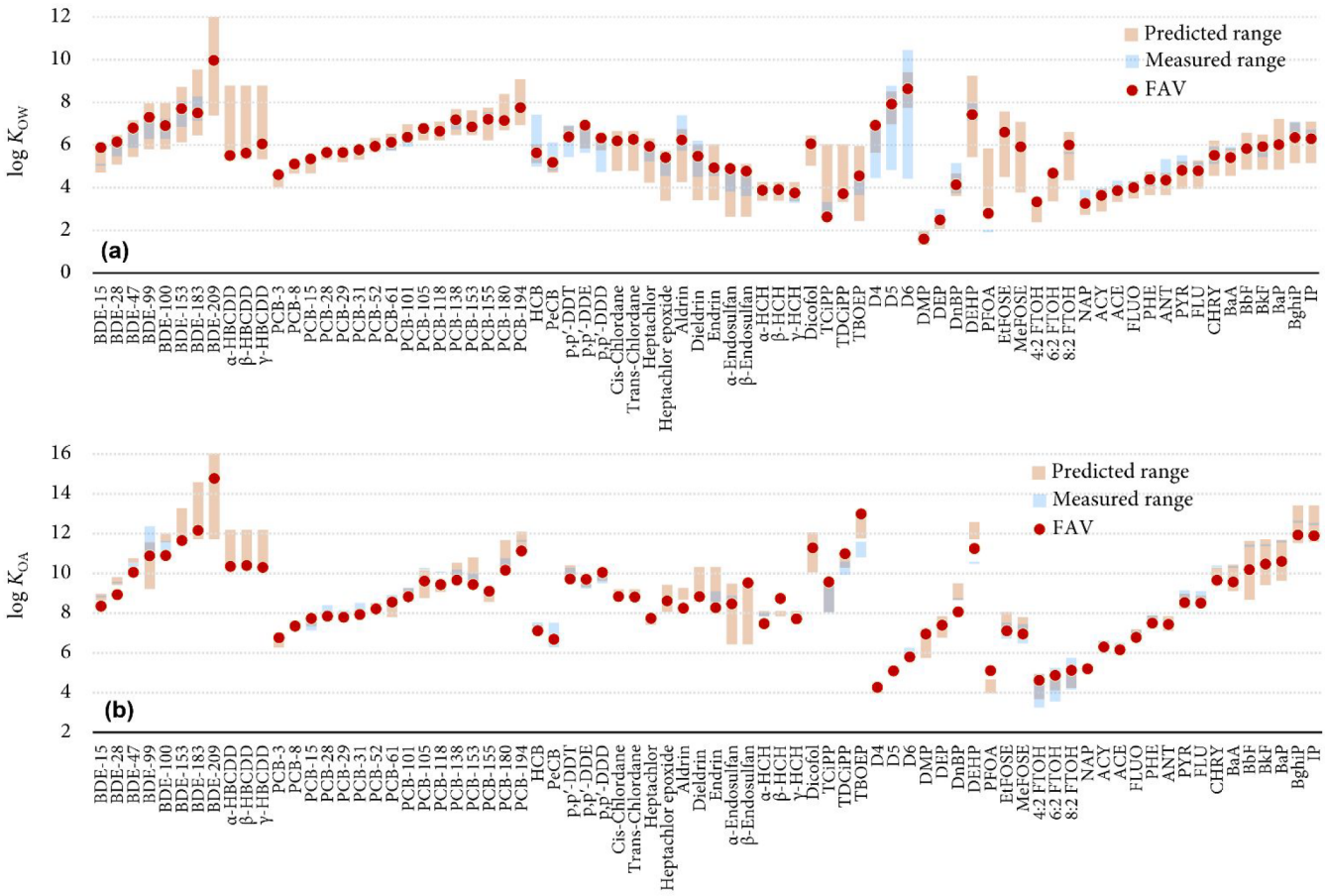


Figure 2. Variability in literature-reported experimentally derived (at a temperature between 20 and 25 °C; blue bars) and in silico predicted (pink bars) K_{OW} (a) and K_{OA} (b) for 76 investigated neutral organic chemicals. The final adjusted values of these chemicals are displayed as red dots for comparison. The variation is missing for some chemicals because only a single laboratory measurement is available from the literature. Here, K_{OW} values are predicted with the KOWWIN module within EPI Suite, OPERA, pp-LFER (with Abraham solute descriptors predicted by IFS-QSAR), MlogP, AlogP, SPARC, and COSMOtherm. K_{OA} values are predicted with the KOAWIN module within EPI Suite, OPERA, pp-LFER (with Abraham solute descriptors predicted by IFS-QSAR), and COSMOtherm. For numeric results, see Supporting Information Table S2.

The literature is full of examples where such empirical correlations feature divergent numerical values. A well-recognized example is the Karickhoff relationship,[110−113] which calculates K_{OC} from K_{OW} and has been reported in many different forms (Table 2). One reason is that the correlations are derived from different empirical data sets, even those with low-quality or erroneous measurements, using different regression settings (for instance, with or without a regression intercept). The earliest form of the Karickhoff relationship, $K_{OC} = 0.411 \times K_{OW}$, was derived from the observed sorption of five polycyclic aromatic hydrocarbons on organic matter-rich sediment,[111] whereas a later, more comprehensive analysis of 117 datapoints of neutral, moderately to highly hydrophobic chemicals ($\log K_{OW} < 7.5$) yielded $K_{OC} = 0.35 \times K_{OW}$.[112] Before using an empirical correlation, one is advised to check the training set for the specific applicable categories of chemicals in terms of hydrophobicity, polarity, planarity, and ionization.[114−116] For example, the Karickhoff relationship may overestimate the sorption of superhydrophobic chemicals because of kinetic limitations to establishing equilibrium partitioning.[112,117] Another reason is that the predicted chemical property is inherently variable under environmentally and biologically relevant conditions. For example, since variability in soil organic matter can cause K_{OC} of the same chemical to vary substantially, it has been recommended to express the Karickhoff relationship as a distribution bounded by the upper and lower limits of $K_{OC} = 0.89 \times K_{OW}$ and $K_{OC} = 0.14 \times K_{OW}$, respectively, rather than discrete numbers derived from one of the many empirical correlations.[112] In this case, the use of $K_{OC} = 0.35 \times K_{OW}$ gives central tendency estimates, with error bounds of a factor of 2.5.[112]

To address the issue of the limited applicability domain of sp-LFERs, environmental chemists attempt to incorporate more descriptors into the empirical correlations. A successful example is pp-LFERs, which express partition ratios and molar enthalpies of phase change as a linear, weighted combination of a series of Abraham solute descriptors, including polarizability/dipolarity parameter (S), solute hydrogen-bond acidity (A), and solute hydrogen-bond basicity (B), as well as McGowan molar volume (V; optional), excess molar refraction (E; optional), or hexadecane−air partition ratio (L; optional).[114,128,129] These solute descriptors are chemical-specific and can be experimentally determined or calculated (e.g., by QSARs/QSPRs, as elaborated below) for most environmental pollutants; the UFZ-LSER database (www.ufz.de/lserd) includes experimentally determined values for over 2000 chemicals,[128] and the ACD/Absolv software includes experimentally determined values for over 5000 chemicals. Each partitioning system has a set of coefficients, called system

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

Table 2. Examples of Empirical Correlations for Predicting the Partitioning in Environmentally or Biologically Relevant Systems from a Chemical's Vapor Pressure, Water Solubility, or Partition Ratios

chemical property	empirical correlation	ref
(I) properties involving pure phases		
vapor pressure (P_L in Pa)	$\log K_{OA} = (-0.988 \pm 0.006), \log P_L + (6.691 \pm 0.017)$ ($N = 222$, $\log K_{OA}$ from −1 to 12, $R^2 = 0.99$)	118
	$K_{OA} \approx 10^{6.7}/P_L$ (the same data set above)	119
water solubility (S_W in mol/L)	the "general solubility equation": $\log K_{OW} = 0.373 - 0.009\,(MP - 25) - 0.978 \times \log S_W$, where MP denotes the melting point (in °C; MP = 25 for MP < 25) ($N = 1026$, $\log K_{OW}$ from −2 to 11, $R^2 = 0.96$)	120
	$K_{OW} \approx 10^{0.42}/S_W$ (the same data set above)	
(II) properties involving environmentally or biologically relevant phases		
organic carbon−water (K_{OC} in L/kg)	the Karickhoff relationship: $K_{OC} = 0.411 \times K_{OW}$ ($N = 5$, $\log K_{OW}$ from 2.12 to 5.18, $R^2 = 0.994$)	111
	$\log K_{OC} = 0.81 \times \log K_{OW} + 0.09$ ($N = 118$, $\log K_{OW}$ from 1.25 to 7.32, $R^2 = 0.89$)	112
	$\log K_{OC} = 0.679 \times \log K_{OW} + 0.663$ ($N = 419$, heterogeneous training set, including 20 acids, $R^2 = 0.912$)	113
phospholipid−water ($K_{plip\text{-}water}$)	$\log K_{nlip\text{-}water} = 1.01 \times \log K_{OW} + 0.12$ ($N = 156$, $\log K_{OW}$ from −2.3 to 8.5, $R^2 = 0.948$)	121
protein−water ($K_{protein\text{-}water}$)	$\log K_{protein\text{-}water} = 0.73 \times \log K_{OW} - 0.39$ ($N = 46$, $\log K_{OW}$ from 1.3 to 6.2, $R^2 = 0.86$)	122
polymeric material−air (K_{PA})	$K_{PA} = 0.06 \times K_{OA}$ ($N = 1273$, $\log K_{OA}$ from 2.0 to 18.7, $R^2 = 0.86$, RMSE = 1.21)	123
polymeric material−water (K_{PW})	$K_{PW} = 0.06 \times K_{OW}$ ($N = 1273$, $\log K_{OW}$ from −0.95 to 9.5, $R^2 = 0.78$, RMSE = 1.21)	123
aerosol−air (K_{PA} in $m^3/\mu g$)	$\log K_{PA} = \log K_{OA} + \log f_m - 11.91$, where f_m is the organic fraction of the aerosol ($N = 59$, $\log K_{OA}$ from approximately 8 to 12.5)	124
dust−air (K_{DA})	$\log K_{DA} = 0.860 \times \log K_{OA} - 0.086$ ($N = 66$, $\log K_{OA}$ from 7.5 to 13.1, $R^2 = 0.78$)	13
carpet−air (K_{CA})	$\log K_{CA} = 0.931 \times \log K_{OA} - 2.408$ ($N = 8$, $\log K_{OA}$ from 1.98 to 4.28)	13
vinyl flooring−air (K_{VA})	$\log K_{VA} = 0.688 \times \log K_{OA} - 0.693$ ($N = 13$, $\log K_{OA}$ from 2.74 to 8.67, $R^2 = 0.80$)	13
cotton−air (K_{CA})	$\log K_{CA} = 0.248 \times \log K_{OA} + 4.447$ ($N = 13$, $\log K_{OA}$ from 7.97 to 12.2, $R^2 = 0.545$)	125
steel−air (K_{SA} in m)	$\log K_{SA} = 0.216 \times \log K_{OA} + 1.420$ ($N = 7$, $\log K_{OA}$ from 10.7 to 14.5, $R^2 = 0.62$)	126
organic film−air (K_{FA})	$\log K_{FA} = 1.100 \times \log K_{OA} - 0.540$ ($N = 62$, $\log K_{OA}$ from 6.8 to 10.6, $R^2 = 0.84 - 0.98$)	127

Table 3. System Constants of Poly-Parameter Linear Free Energy Relationships (pp-LFERs) for Calculating Partition Ratios and Enthalpies[a]

	e	s	a	b	v	l	constant	ref
(I) properties involving pure phases								
octanol−air partition ratio (K_{OA}) at 25 °C		0.42	3.52	0.84		0.92	−0.12	131
octanol−water partition ratio (K_{OW}) at 25 °C		−1.41	−0.18	−3.45	2.41	0.43	0.34	26
air−water partition ratio (K_{AW}) at 25 °C		−2.07	−3.67	−4.87	2.55	−0.48	0.59	132
octanol−air transfer enthalpy (ΔH_{OA} in kJ/mol)		6.04	−53.66	−9.19	1.57	−9.66	−6.67	133
octanol−water transfer enthalpy (ΔH_{OW} in kJ/mol)		5.31	−20.1	34.27	18.88	−8.26	1.74	b
air−water transfer enthalpy (ΔH_{AW} in kJ/mol)		−0.73	33.56	43.46	17.31	1.40	8.41	133
(II) properties involving environmentally or biologically relevant phases								
organic carbon−water (K_{OC} in L/kg) at 25 °C		−0.98	−0.42	−3.34	1.20	0.54	0.02	134
storage of lipid−water ($K_{nlip\text{-}water}$) at 37 °C	0.70	−1.08	−1.72	−4.14	4.11		−0.07	135
phospholipid−water ($K_{plip\text{-}water}$) at 37 °C	0.74	−0.72	0.11	−3.63	3.30		0.29	121
protein−water ($K_{protein\text{-}water}$) at 37 °C	0.51	−0.51	0.26	−2.98	3.01		−0.65	122
polymeric material−air (K_{PA}) at 25 °C		−0.1	2.4	−0.4	−2.3	1.4	−0.2	123
polymeric material−water (K_{PW}) at 25 °C		−0.9	−2.7	−3.9	0.8	0.7	−0.9	123

[a]The pp-LFERs share the general form, $\log K = s \times S + a \times A + b \times B + e \times E + v \times V + l \times L + \text{constant}$. [b]Calculated from ΔH_{OA} and ΔH_{AW} in this table.

constants, that is calculated by calibration against experimentally derived data in a training set.

Table 3 aggregates published pp-LFERs for partition ratios broadly used in chemical assessments. The UFZ-LSER database includes a more comprehensive list of pp-LFERs. Compared to sp-LFERs, pp-LFERs have several advantages for use in chemical assessment:[114,128,129] (1) pp-LFERs are not limited to certain categories of chemicals and have a wider applicability domain, but they are not any more challenging to use than sp-LFER; (2) predictions from pp-LFERs have been demonstrated to agree well with experimentally determined data; and (3) the combination of each solute descriptor and the associated system coefficients mechanistically reflects a certain intermolecular interaction between chemical and the solvating phases, and its magnitude indicates the relative importance of this interaction in the overall partitioning. For a more wholesome discussion of pp-LFERs, we refer to the reviews by Goss and colleagues.[114,128,129] It should also be noted that pp-LFERs may be evolving as more diverse, heterogeneous chemical categories become incorporated in the calibration set. Various versions of pp-LFER equations are available in the literature, and some versions may be

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

constrained if certain chemical categories, such as PFASs and organosilicons, are missing from the training set.[130]

QSARs/QSPRs

Several QSAR/QSPR prediction tools are currently available for free for predicting chemical properties commonly used in chemical assessment, including the Abraham solute descriptors used for parametrizing pp-LFERs (Table 4), such as the Estimation Programs Interface (EPI) Suite,[136] OPEn structure−activity/property Relationship App (OPERA),[137] QSAR-INSubria-Chem (QSARINS-Chem),[138,139] and Iterative Fragment Selection (IFS-QSAR).[140−143] In addition to these free prediction tools, proprietary software, such as ACDLabs and SPARC, can be used to make predictions. While all these tools predict chemical properties based on the extrapolation or interpolation of experimentally derived data for chemicals in a training set, they utilize different algorithms for extrapolation or interpolation. For example, the QSARs/QSPRs in EPI Suite are largely atom/fragment contribution methods,[144] which segment a molecular structure into atoms or structural fragments, sum the products of predefined constant coefficients for individual atoms/fragments with the number of occurrences of the atoms/fragments in the molecule to obtain an estimate of the property value, which, if necessary, is corrected for certain preidentified structural arrangements. The QSARs/QSPRs developed with the IFS algorithm also rely on the fragmentation of a molecular structure, but this algorithm automatically selects a minimum subset of the most influential fragments and combines them as the final estimate of the property value.[140−142] Instead of using atoms and structural fragments, QSARs/QSPRs in QSARINS-Chem first calculate a series of more holistic molecular descriptors (e.g., the states of mono- and/or bidimensional conformations) and fingerprints, select the most statistically relevant descriptors using the genetic algorithm, and then linearly combine their values as the final estimate of the property value, with coefficients calibrated from the training set using multivariate linear regression.[138,139] Like QSARINS-Chem, OPERA also starts with calculating molecular descriptors, but it selects the five chemicals that are most structurally similar to the chemical of interest ("the five nearest neighbors") from the training set and predicts the property value by averaging the experimentally determined values of these five chemicals weighted by their proximity (the distance in the chemical space).[137]

Table 4. Calculation of Chemical Properties by Commonly Used Nonproprietary QSAR/QSPR Tools ("Y" = Predicted; "N" = Not Predicted)

	EPI Suite	OPERA	IFS-QSAR	QSARINS-Chem
output of applicability domain information	N[a]	Y	Y	Y
Abraham solute descriptors	N	N	Y	Y
chemical properties				
boiling point	Y	Y	Y	N
melting point	Y	Y	Y	N
K_{OW}, K_{OA}, K_{AW}	Y	Y	Y[b]	N
K_{OC}	Y	Y	Y[b]	Y
vapor pressure	Y	Y	Y	N
water solubility	Y	Y	Y	N
dissociation constants (pK_a or pK_b)	N	Y	N	N
reaction with the hydroxy radical	Y	Y	N	Y
reaction with ozone	Y	N	N	Y
biodegradation	Y[c]	Y	N	N
hydrolysis	Y[d]	N	N	N
biotransformation in fish	Y	Y	Y	Y
biotransformation in humans	N	N	Y	Y

[a]EPI Suite does not provide explicit information on the applicability domain. Instead, EPI Suite documentation provides general guidance to identify "less reliable predictions", including chemicals that (i) are outside the experimental range of the training set, (ii) are outside the range of molar mass of chemicals in the training set, (iii) have more instances of a given fragment than the maximum among all chemicals in the training set, and/or (iv) contain structural features not represented in the training set.[136] One can develop an "in-house" method to determine whether a prediction is within the applicability domain.[8] [b]IFS-QSAR can predict Abraham solute descriptors (without discriminating between isomers), which can be combined with system constants for the prediction of partition ratios. One can also use experimentally derived Abraham solute descriptors to predict partition ratios; in this case, if isomer-specific Abraham solute descriptors are used, then the pp-LFER can provide isomer-specific predictions. [c]EPI Suite calculates the probability of primary biodegradation, which can be extrapolated to the biodegradation half-life using an empirical relationship by Arnot et al.[152] [d]Only acid- and base-catalyzed hydrolysis are considered; the prediction tool does not estimate neutral hydrolysis rate constants.

One needs to exercise caution when using and interpreting QSAR/QSPR predictions. First, since not all QSARs/QSPRs are of the same quality, their predictions are not equally credible or reliable. Preference should be given to QSARs/QSPRs that have been well-validated based on the *OECD Guidance Document on the Validation of (Q)SAR Models*[145] and provide detailed documentation of key information in compliance with the QSAR Model Reporting Format. Second, one must ensure that the QSAR/QSPR prediction matches exactly the property required by the chemical assessment, that is, the QSAR/QSPR is "fit-for-purpose".[146] For instance, since most QSARs/QSPRs do not consider chemical dissociation and stereochemistry, the predictions can only be used as values for the neutral form of a chemical without discrimination between stereoisomers. Likewise, since most chemical assessment models require half-lives for primary biodegradation (that is, the disappearance of the parent compound), one needs to avoid the misuse of the QSAR/QSPR predictions for ultimate biodegradation (that is, the complete breakdown of all organic carbons into carbon dioxide or methane).

Third, it is important to know whether a prediction falls within the applicability domain (AD) of a QSAR/QSPR.[146] The AD describes the coverage of the training set, or the structural and property space, in which a QSAR/QSPR makes predictions by interpolation rather than extrapolation.[145,147] The determination of whether a chemical of interest falls within the AD of a QSAR/QSPR is based on the similarity between this chemical and the chemicals in the training set. A chemical being within the AD indicates the compliance of the QSAR/QSPR's specifications with the intended use case, and therefore, the use of the QSAR/QSPR is legitimate and suitable.[147] Statistically speaking, restricting the use of a QSAR/QSPR to its AD generally improves its predictive performance.[23] However, a chemical being within the AD does not necessarily indicate or guarantee the reliability or accuracy of every single prediction. For instance, the "five nearest neighbors" algorithm by OPERA predicts a chemical property as the weighted average of the experimentally determined data

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

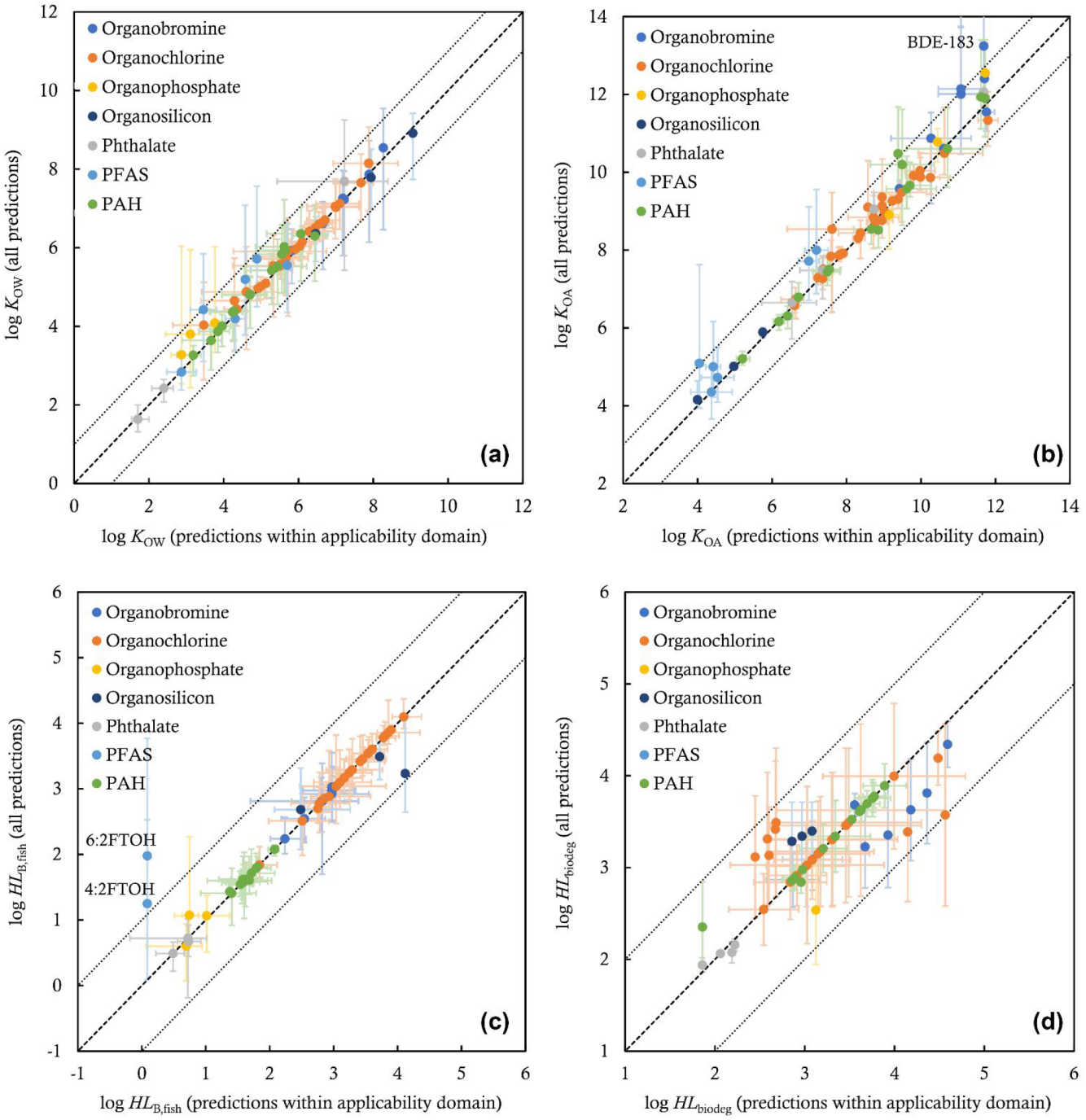


Figure 3. Consensus values (arithmetic means of log-transformed values) of QSAR/QSPR predictions of octanol−water (K_{OW}) and octanol−air (K_{OA}) partition ratios, fish biotransformation half-life ($HL_{B,fish}$; normalized to a body weight of 10 g), and primary biodegradation half-life (HL_{biodeg}), with and without excluding predictions outside the applicability domains (ADs) of the corresponding QSARs/QSPRs. Diagonal dashed lines indicate perfect agreement between two sets of consensus values, and dotted lines indicate an order-of-magnitude discrepancy between two sets of consensus values. Here, K_{OW} values are predicted with the KOWWIN module within EPI Suite, OPERA, pp-LFER (Abraham solute descriptors predicted by IFS-QSAR), MlogP, AlogP, SPARC, and COSMOtherm. K_{OA} values are predicted with the KOAWIN module within EPI Suite, OPERA, pp-LFER (Abraham solute descriptors predicted by IFS-QSAR), and COSMOtherm. $HL_{B,fish}$ values are predicted with OPERA, IFS-QSAR, and QSARINS. HL_{biodeg} values are predicted with EPI Suite (half-lives converted from the predicted primary biodegradation probabilities according to Arnot et al.[152]) and OPERA. Since no AD information is available for EPI Suite, we develop an "in-house" method to determine whether a prediction is within the AD of EPI Suite based on the description in the EPI Suite technical documentation (see the footnote of Table 4). Since no AD information is available for SPARC and COSMOtherm, we do not consider them to be within the AD. In addition, a pp-LFER prediction is deemed to be within the AD if all the predicted Abraham solute descriptors are within the ADs of both the pp-LFER equations[156] and IFS-QSAR.[143]

of the five most structurally similar chemicals in the training set. Such an averaging operation prevents any predictions from exceeding the upper bound of the training set. For chemicals structurally similar to chemicals constituting the upper bound

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

of the training set but with their "true" values beyond the upper bound, the algorithm would identify them as being within the AD but underpredict their values. A case is BDE-183 (with a log K_{OA} FAV of 12.16; Figure 2): while it is identified as being within the AD because it is structurally similar to five chemicals in the OPERA training set, OPERA underestimates its log K_{OA} (11.69) because the FAV exceeds the maximum log K_{OA} (defined by octachlorodibenzo-p-dioxin; log K_{OA} = 12.05) among its five nearest neighbors, and even the maximum log K_{OA} in the entire training set (1-eicosanol; log K_{OA} = 12.06). In fact, 1-eicosanol sets a cap of approximately 12 for all of OPERA's log K_{OA} predictions. This case demonstrates that predictions within the AD should not be unconditionally trusted. Conversely, a prediction may be located outside the AD but can still be reliable or accurate.[145] Furthermore, the definition of AD differs among QSARs/QSPRs and depends closely on the data structure of the training set.

Lastly, one needs to recognize the uncertainty associated with experimentally derived values in the training set. The quality of the experimentally derived data may also be questionable because of errors in data selection and curation (see a list of common errors summarized by Dearden et al.[148]). Overall, a QSAR/QSPR's mathematical extrapolation or interpolation approach adds uncertainty to the predictions beyond the uncertainty inherent in the training set.[149] To address this limitation, property prediction techniques based on quantum chemistry and statistical thermodynamics, e.g., COSMOtherm, serve as an ideal complement because they rarely, or to a very limited extent, rely on experimentally derived data. Recent studies have demonstrated that COSMOtherm performed almost as well as QSARs/QSPRs based on training sets in predicting partition ratios and internal energies, with the exception of the internal energy change for dissolution in water.[131,150,151] Nevertheless, COSMOtherm calculations usually require a long time and intensive computational power because they rely on 3D geometry optimization of the molecular structure, which limits the use of this approach in high-throughput chemical screening.

When multiple QSAR/QSPR predictions are available for a single property, the mean (e.g., geometric, arithmetic, harmonic, or least-squares mean) or other central-tendency estimates, referred to as the "consensus value", are recommended as a reasonable estimate to combine the battery of QSAR/QSPR predictions for chemical assessments.[23,153,154] Such a practice rests on the assumption that QSARs/QSPRs building on different algorithms may consider, or assign different weights to, different particular molecular structures or descriptors and that errant predictions can, therefore, be mitigated by predictions from other models.[155] A systematic comparison between 15 QSAR models for aqueous toxicity indicated superior predictivity of consensus values when compared to individual models.[23] Notably, the calculated consensus values were fairly consistent regardless of whether or not predictions outside the AD of the QSARs were excluded from the calculation. Those authors, therefore, concluded that the use of AD was "less important" for consensus modeling, at least for the specific data set used in their work, and called for more tests on additional data sets.[23]

As an example, Figure 3 compares the consensus values (arithmetic means) of log K_{OW}, log K_{OA}, log $HL_{B,fish}$, and log HL_{biodeg} predictions for the 76 investigated neutral organic chemicals, with and without considering the coverage of each prediction by the AD of each QSAR/QSPR. For chemicals falling within the AD of at least one QSARs/QSPRs, the two sets of consensus values agree generally well, with the discrepancy within an order of magnitude. As such, it makes little difference for the calculation of consensus values whether AD is considered or not. An exception is the calculated consensus value of log K_{OA} for BDE-183, where the consideration of AD generates a value (11.69) that is 1.66 log units lower than that obtained when disregarding AD (13.34). This is because BDE-183 only falls within the AD of OPERA, whose predictions are capped at approximately 12, as discussed above. Likewise, the calculated consensus values of log $HL_{B,fish}$ for 4:2 FTOH and 6:2 FTOH are 1.2 and 1.8 log units lower if AD is considered, as these two chemicals fall only within the AD of the IFS-QSAR, whose predictions ($HL_{B,fish}$ = 1.2 h for both chemicals) are much shorter than predictions by OPERA (3 and 14 h, respectively) and QSARINS-Chem (76 and 763 h, respectively). Overall, the consideration of AD may lead the consensus value to skew toward a single model's prediction. Interestingly, if comparing the log K_{OA} consensus values with the FAV of BDE-183, one can conclude that the consideration of AD gives slightly better agreement with the FAV (12.16; Figure 2). However, no measurement is available for evaluating whether this is also the case for log $HL_{B,fish}$. We cannot exclude the possibility that the only working QSAR/QSPR may give unsatisfactory predictions because the AD does not guarantee the accuracy of the prediction. Also, the consideration of AD precludes hexabromocyclododecane (HBCDD) isomers, PFASs, and certain pesticides and organophosphates from the calculation of consensus values of log K_{OA}, log $HL_{B,fish}$, and log HL_{biodeg}, respectively, because they fall outside the ADs of all QSARs/QSPRs used here.

Integrated Online Platforms

With the development of cloud storage and computing, integrated online platforms become available for efficiently obtaining chemical property data. Examples of these integrated online platforms include PubChem (developed and maintained by the U.S. National Institute of Health), ChemSpider (developed and maintained by the U.K. Royal Society of Chemistry), CompTox (developed and maintained by the U.S. Environmental Protection Agency), ChemIDplus (developed and maintained by the U.S. National Library of Medicine), eChemPortal (developed and maintained by OECD), the REACH registered substances database (developed and maintained by the ECHA), and EAS-E Suite (developed and maintained by ARC Arnot Research & Consulting). In addition to these integrated online platforms, SciFinder, developed and maintained by the Chemical Abstracts Service, is also a useful online tool for mining chemical property data based on its huge number of chemical and bibliographic records.

The main advantage of these online platforms is that they store, organize, and integrate chemical data (e.g., physical–chemical properties, environmental fate, biomonitoring, toxicokinetic data) for thousands of compounds in online databases that users can query and retrieve desired information. The platforms are generally provided with an interactive graphical user interface that helps users in browsing and searching data. In general, compounds can be searched by the CAS Registration Number and/or name (IUPAC and common names). Some platforms allow the use of other identifiers such as DSSTox substance identifier (DTXSID), or

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

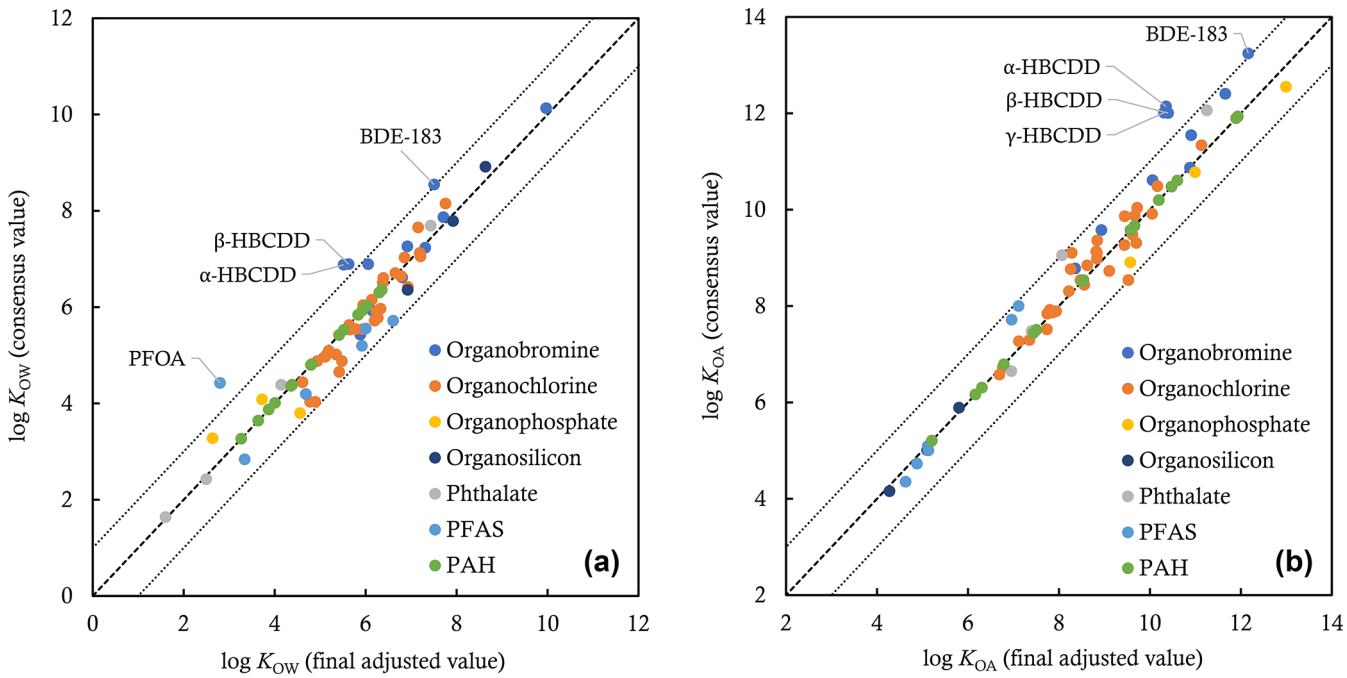


Figure 4. Comparison between the final adjusted values of laboratory measurements and consensus values (arithmetic means of log-transformed values) of in silico predictions (without excluding predictions outside corresponding ADs) for the 76 investigated neutral organic chemicals. Chemicals with a difference of greater than 1 order of magnitude are identified by their names. Diagonal dashed lines and dotted lines indicate perfect agreement and an order-of-magnitude discrepancy, respectively, between FAVs and in silico consensus values. See the caption of Figure 2 for in silico prediction tools used to calculate K_{OW} and K_{OA}.

structural information such as the International Chemical Identifier (InChI) Key, and Simplified Molecular Input Line Entry System (SMILES).[157] These online platforms store compiled and curated experimentally derived data, which are collected from multiple sources through aggregation and standardization. For instance, PubChem integrates measured chemical property data in the Hazardous Substance Data Bank, ILO International Chemical Safety Cards, and the Occupational Safety and Health Administration data, whereas CompTox integrates chemical property data in the PhysProp data set with environmental fate, exposure and toxicokinetic data. The online platforms also contain empirical correlations and QSARs/QSPRs to allow computation of chemical property data, allowing users to make predictions for chemicals that are not immediately included in the curated databases. For example, CompTox integrates predictions of OPERA, EPI Suite, ACD/Laboratories, and Toxicity Estimation Software Tool, and EAS-E Suite contains predictions of EPI Suite, OPERA, IFS-QSAR, and QSARINS-Chem. Therefore, these online platforms link experimentally derived data and in silico predictions and, therefore, aggregate and connect scientific information from different fields. This is relevant in the context of risk assessment and chemical evaluation as these platforms represent places where curated data and metadata can be easily retrieved and compared.

Moreover, these online platforms include curated structural information such as InchIKey and/or SMILES that can be downloaded and used to obtain predictions from QSPRs or QSARs. For instance, for each chemical, CompTox contains a standardized "QSAR-ready" SMILES string subject to "kekulization" (i.e., the localization of double bonds in aromatic rings),[158] whereas EAS-E Suite includes a standardized canonical SMILES string that is compatible with the IFS-QSAR input requirement.

Another main advantage of the online integrated platforms is that they are actively maintained and periodically updated. The integrated online platforms are of value for scientific and regulatory communities; they also benefit from these communities because users can help improve the system by submitting new data and reporting bugs or errors.

It needs to be kept in mind that these online platforms may contain inaccuracies or errors since they are secondary sources aggregated and curated from other databases and publications.[159−162] Examples include mismatched chemical identifiers (e.g., a CAS Registry Number or name is incorrectly assigned to one or more irrelevant structures), incorrectly reported chemical structures (e.g., omitting stereochemistry or tautomerization), duplicated data, and uninterpretable data due to a lack of additional relevant information. Efforts to address such inaccuracies and errors are already in place, and many online platforms have implemented systems to filter and validate their data. For instance, ChemSpider has established a daily curation feed to summarize the validation and deletion operations of name-structure relationships.[159] Other data curation workflows have been developed to check and verify chemical information automatically retrieved from the databases.[158,163]

VARIABILITY AND UNCERTAINTY IN CHEMICAL PROPERTIES

Both experimentally derived data and in silico predictions can be retrieved from multiple sources, and therefore, the values may cover wide ranges. This is even the case for properties related to the air−water−octanol system, which can be measured using test standards and guidelines and, therefore, should theoretically have single "true" values. With the illustrative cases of partition ratios, Figure 2 compares the

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

FAVs of the log K_{OW} and log K_{OA} of the 76 investigated neutral organic chemicals against the ranges of the collected experimentally derived data (without excluding data identified to be erroneous, unreliable, or of low quality and excluded from the calculation of FAVs) and the ranges of predictions by the in silico tools used in this paper (without excluding predictions outside the ADs). The range is particularly large among log K_{OW} measurements for organosilicons (Figure 2a) and among log K_{OA} measurements for organophosphates and fluorotelomers (Figure 2b). The great variability in experimentally derived data results mainly from differences in measurement technique, different data quality and reliability, as well as inter- and intra-laboratory variations, as elaborated in earlier sections. On the other hand, in silico predictions are also subject to the variability of nearly 2 orders of magnitude (Figure 2) because different empirical correlations and QSARs/QSPRs build on different training sets and adopt distinct algorithms. The discrepancy between predictions is the largest for chemicals with little overlap with the ADs of the QSARs/QSPRs, such as highly brominated diphenyl ethers, HBCDD isomers, endosulfan isomers, and organosilicons (Figure 2).

For the 76 investigated neutral organic chemicals (Table S3), OPERA is the best at reproducing the log K_{OW} FAVs (R^2 = 0.902; root-mean-square deviation or RMSE = 0.478 units), followed by COSMOtherm (R^2 = 0.865; RMSE = 0.557 units) and the KOWWIN module in EPI Suite (R^2 = 0.846; RMSE = 0.757 units). The pp-LFER with solute descriptors predicted by IFS-QSAR is best at reproducing the log K_{OA} FAVs (R^2 = 0.931; RMSE = 0.784 units), compared to the KOAWIN module in EPI Suite (R^2 = 0.879; RMSE = 0.843 units) and OPERA (R^2 = 0.873; RMSE = 0.780 units). These comparisons indicate that certain laboratory measurements or in silico predictions may depart far from the FAVs. As such, arbitrary selection of laboratory measurements or in silico predictions, without a comprehensive examination of their quality and consistency, may bring about substantial biases in chemical assessments.

As elaborated earlier, the FAVs represent harmonized central tendencies of experimentally derived data, and the consensus values represent harmonized central tendencies of in silico predictions regardless of the exclusion of predictions outside of the ADs. Figure 4 displays that the FAVs agree well with the consensus values (arithmetic means of log-transformed values in this case), with a discrepancy within an order of magnitude, for most of the 76 neutral organic chemicals investigated here. Statistically, the consensus values agree better with the FAVs for both log K_{OW} (R^2 = 0.907 and RMSE = 0.472 units) and log K_{OA} (R^2 = 0.935 and RMSE = 0.573 units) than any of the predictions by individual QSARs/QSPRs (Table S3). Interestingly, the combined use of the two best performing models (OPERA and COSMOtherm), instead of all, gives even better performance for predicting log K_{OW} (R^2 = 0.930 and RMSE = 0.404 units), whereas the combined use of the three best performing models (pp-LFER, EPI Suite, and OPERA) give slightly better performance for predicting log K_{OA} (R^2 = 0.946 and RMSE = 0.531 units). Therefore, it is not necessary to use all available in silico prediction tools in the calculation of consensus values. As Table S4 shows, such agreement is apparent not only for the "traditional" nonpolar, semivolatile chemicals such as organochlorines (RMSE of 0.357 and 0.359 units for log K_{OW} and log K_{OA}, respectively, N = 34) and PAHs (RMSE of 0.190 and 0.560 units for log K_{OW} and log K_{OA}, respectively, N = 15) but also for "emerging" chemicals such as the polar neutral PFASs (RMSE of 0.881 and 0.495 units for log K_{OW} and log K_{OA}, respectively, N = 6) and the highly volatile organosilicons (RMSE of 0.374 and 0.099 units for log K_{OW} and log K_{OA}, respectively, N = 3). Hence, we believe that the use of consensus values of in silico predictions in chemical assessment models is justified and defensible, if experimentally derived data are missing, inadequate, flawed, or lacking reported uncertainty information for the chemical of interest. Such a conclusion is important because the use of in silico predictions is often unavoidable for premanufactured chemicals or chemicals in design and for model-based high-throughput chemical screening and prioritization.[7,8] In addition, Table S5 shows that arithmetic and harmonic means of log-transformed values show similar agreement with the FAVs, and the arithmetic means of nontransformed values perform less satisfactorily in reproducing the FAVs because they tend to be biased by high extreme values. The use of different types of means (e.g., geometric, arithmetic, or harmonic) in calculating the consensus values of different chemical properties also warrants considerations in the statistical distribution of in silico predictions,[164] the nature of the properties,[44,164] and the intended context of chemical assessments (e.g., whether conservativeness is desired).[165]

Previously, Puzyn[149] found that for 12 organochlorines and hydrocarbons, the *OECD POV and LRTP Screening Tool* makes consistent predictions for their potential for persistence and long-range transport, regardless of whether the model was parametrized with experimentally derived or QSAR/QSPR-predicted partition ratios and half-lives. The author, therefore, "strongly recommend[ed] revising the OECD recommendations that give the highest priority to the application of purely empirical data in environmental mass balance modeling".[149] Likewise, Tebes-Stevens et al.[154] recommended that "without adequate data to perform a rigorous comparison of individual calculator performance against measured data, the geometric mean and the median of the predicted values from multiple calculators provide more robust estimates of the property value than any individual calculator."

Figure 4 also displays exceptions. For K_{OW}, the consensus values deviate from the FAVs by 1.36 and 1.26 log units for α- and β-HBCDDs, 1.04 log units for BDE-183, and 1.62 log units for the neutral form of PFOA. For K_{OA}, the consensus values deviate from the FAVs by 1.60 to 1.79 log units for HBCDD isomers, and 1.08 log units for BDE-183. Here, HBCDDs and BDE-183 fall outside the ADs of almost all of the QSARs/QSPRs used here. K_{OW} predictions for the neutral form of PFOA fall within the ADs of OPERA (predicted log K_{OW} = 3.11) and AlogP (predicted log K_{OW} = 3.81); these values are closer to the FAV (2.80) compared to their counterparts outside the AD (5.21 by pp-LFER, 4.81 by EPI Suite, and 4.17 by MlogP). However, the opposite is seen for β-endosulfan: although K_{OA} predictions are located in the ADs of OPERA (predicted log K_{OA} = 8.81) and EPI Suite (predicted log K_{OA} = 6.41), the pp-LFER prediction that is outside the AD (9.45) performs much better in reproducing the FAV (9.53). Due to the small sample size, we cannot arrive at a general conclusion as to whether the exclusion of in silico predictions outside of the AD from the consensus estimate can give consensus values that agree better with the FAVs than the consensus estimate without such exclusions.

As originally conceived, LDVs, and therefore also the FAVs derived from them, are exclusively based on experimentally

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395

derived data.[9,94,95,99,100,166,167] However, more recent efforts reinterpreted the definition of LDVs by incorporating and tuning in silico predictions with laboratory measurements.[150,168] We disagree with this reinterpretation of the LDV/FAV concept for primarily two reasons. First, since many in silico prediction tools are developed through calibration against laboratory measurements, the results of these prediction tools inherit the uncertainty in the laboratory measurements and are not necessarily independent of laboratory measurements. Notably, for chemicals in the training sets of in silico prediction tools, the in silico predictions may even be close or identical to the laboratory measurements used to construct the in silico tools, which is especially the case of tools based on the "nearest neighbors" method such as OPERA. As such, including these in silico predictions duplicates the laboratory measurements in the calculation of LDVs and, somewhat, biases the calculated LDVs toward these laboratory measurements. Second, while many in silico predictions are internally consistent, such an internal consistency is ostensible because the tools actually adopt the internal consistency assumption, rather than independent algorithms, to predict properties from other predicted properties. An example is that EPI Suite calculates K_{OA} as the ratio of independent predictions of K_{OW} and K_{AW}. Such K_{OA} predictions do not add independent information to the adjustment of FAVs, and moreover, if K_{OW} and K_{AW} predictions are biased, then the biases will be propagated to make K_{OA} predictions highly uncertain. In this paper, we therefore advocate for, and adopt, the original definition of the LDVs as being solely based on experimental data.

RECOMMENDATIONS FOR CHEMICAL ASSESSMENTS

Both, experimentally derived and in silico predicted chemical properties, are subject to considerable uncertainty, which is readily apparent from the variability in the values obtained from different studies. Importantly, not all experimentally derived data share the same quality and reliability, and therefore, they are not always superior to in silico predictions. On the other hand, in silico derived values may be limited by the applicability and predictivity of the prediction tools. The use of randomly or arbitrarily selected experimentally derived data or in silico predictions can introduce significant uncertainty and even errors to the assessments of chemical emissions, fate, hazard, exposure, and risks.

All things considered, for reliable and defensible chemical assessments, we recommend using either carefully selected LDVs (or FAVs), if a sufficient number of reliable laboratory measurements of a chemical's property is available, or the consensus values of multiple in silico prediction tools, if the data pool from laboratory measurements is not adequate and if there is no particular reason to believe that one prediction method is better than the others. This is because a single laboratory measurement may suffer from unknown uncertainty, inconsistency, bias, and even error, and on the other hand, different in silico tools build on distinct training sets, employ different prediction algorithms, and therefore demonstrate suitability for different parts of the chemical space. Harmonization of multiple experimentally derived data or in silico predictions is most likely to give reasonable estimates for chemical properties. In cases where only a single experimental measurement is available, it can be used for chemical assessments if its quality is evaluated to meet the requirements of quality control and assurance. Also, the LDVs (or FAVs) can be used as standards to examine the predictivity of different in silico prediction tools, whereas the consensus values can also guide the selection of laboratory measurements to derive the most reasonable LDVs or FAVs. The evaluation of chemical properties is not necessarily a linear procedure; rather, it requires the iteration between laboratory measurements and in silico predictions. Table S6 shows different scenarios of how harmonized FAVs respond to the selection of laboratory measurements for deriving LDVs and how the consensus values of multiple in silico prediction tools inform the plausibility of the harmonized FAVs. Such an iterative procedure cannot be achieved without individual precise, accurate laboratory measurements and the constant improvement and development of in silico prediction tools.

To date, while the use of LDVs (or FAVs) has been recognized, recommended, and practiced by a series of model-based chemical assessments,[9,96,100,167] the use of consensus values of in silico predictions is still limited. Recently, Li et al.[8] developed a workflow of auto-parameterization of chemical emission, fate, exposure, and risk models using the consensus values of QSARs/QSPRs. The auto-parameterized model succeeded in reproducing the inferred U.S. general population exposure estimates, with 79 and 97% of the predictions deviating from the observation-based inferences by a factor of 10 and 100.[8] This case demonstrates the appealing application of consensus values in chemical assessments, which is particularly useful for high-throughput screening and prioritization of "data-poor" chemicals. Such a workflow has also been implemented in EAS-E Suite, allowing users to auto-parameterize built-in models with the same set of standardized, coherent inputs.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsenvironau.2c00010.

Tables documenting the literature-derived values of 16 neutral organic chemicals; tables comparing the performance of QSARs/QSPRs in reproducing the final adjusted values of 76 neutral organic chemicals; table comparing the final adjusted values based on the selection of different sets of laboratory measurements- (PDF)

Tables documenting the final adjusted values and consensus values of 76 neutral organic chemicals (XLSX)

AUTHOR INFORMATION

Corresponding Author

Li Li − *School of Public Health, University of Nevada Reno, Reno, Nevada 89557, United States;* orcid.org/0000-0002-5157-7366; Phone: +1 (775) 682 7077; Email: lili@unr.edu

Authors

Zhizhen Zhang − *School of Public Health, University of Nevada Reno, Reno, Nevada 89557, United States*

Yujie Men − *Department of Chemical & Environmental Engineering, University of California Riverside, Riverside, California 92521, United States;* orcid.org/0000-0001-9811-3828

Sivani Baskaran − *Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario M1C 1A4, Canada*

Alessandro Sangion − *Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario M1C 1A4, Canada; ARC Arnot Research & Consulting, Toronto, Ontario M4M 1W4, Canada*

Shenghong Wang − *School of Public Health, University of Nevada Reno, Reno, Nevada 89557, United States*

Jon A. Arnot − *Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario M1C 1A4, Canada; ARC Arnot Research & Consulting, Toronto, Ontario M4M 1W4, Canada; Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario M5S 1A8, Canada;* orcid.org/0000-0002-4295-4270

Frank Wania − *Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario M1C 1A4, Canada;* orcid.org/0000-0003-3836-0901

Complete contact information is available at: https://pubs.acs.org/10.1021/acsenvironau.2c00010

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

L.L., Z.Z., A.S., and J.A.A. acknowledge funding from the European Chemistry Industry Council Long-range Research Initiative (CEFIC-LRI; ECO54). This publication has not been formally reviewed by the funding agency, and the views expressed in this publication are solely those of the authors. We thank Peter Fantke for his insightful discussions on the derivation of consensus values of in silico predictions with different types of means in statistics.

REFERENCES

(1) European Parliament. *Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH)*; 2006; https://osha.europa.eu/en/legislation/directives/regulation-ec-no-1907-2006-of-the-european-parliament-and-of-the-council (accessed 2022-07-12).

(2) United Nations Environment Programme. *Stockholm Convention on Persistent Organic Pollutants (Text and Annexes)*, 2001; http://chm.pops.int (accessed 2022-07-12).

(3) Arnot, J. A.; Gobas, F. A. A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. *Environ. Rev.* **2006**, *14* (4), 257−297.

(4) Scheringer, M.; Jones, K. C.; Matthies, M.; Simonich, S.; Van De Meent, D. Multimedia partitioning, overall persistence, and long-range transport potential in the context of POPs and PBT chemical assessments. *Integr. Environ. Assess. Manag.* **2009**, *5* (4), 557−576.

(5) Neumann, M.; Schliebner, I. *Protecting the Sources of Our Drinking Water: The Criteria for Identifying Persistent, Mobile and Toxic (PMT) Substances and Very Persistent and Very Mobile (vPvM) Substances under EU Regulation REACH (EC) No 1907/2006*; German Environment Agency (Umweltbundesamt): Dessau-Roßlau, 2019.

(6) Zhang, X.; Sühring, R.; Serodio, D.; Bonnell, M.; Sundin, N.; Diamond, M. L. Novel flame retardants: Estimating the physical−chemical properties and environmental fate of 94 halogenated and organophosphate PBDE replacements. *Chemosphere* **2016**, *144*, 2401−2407.

(7) Arnot, J. A.; Brown, T. N.; Wania, F.; Breivik, K.; McLachlan, M. S. Prioritizing chemicals and data requirements for screening-level exposure and risk assessment. *Environ Health Perspect* **2012**, *120* (11), 1565−1270.

(8) Li, L.; Sangion, A.; Wania, F.; Armitage, J. M.; Toose, L.; Hughes, L.; Arnot, J. A. Development and evaluation of a holistic and mechanistic modeling framework for chemical emissions, fate, exposure, and risk. *Environ. Health Perspect.* **2021**, *129*, 127006.

(9) Schenker, U.; MacLeod, M.; Scheringer, M.; Hungerbühler, K. Improving data quality for environmental fate models: A least-squares adjustment procedure for harmonizing physicochemical properties of organic compounds. *Environ. Sci. Technol.* **2005**, *39* (21), 8434−8441.

(10) Meyer, T.; Wania, F.; Breivik, K. Illustrating sensitivity and uncertainty in environmental fate models using partitioning maps. *Environ. Sci. Technol.* **2005**, *39* (9), 3186−3196.

(11) European Chemicals Agency. *Guidance on Information Requirements and Chemical Safety Assessment. Part C: PBT/vPvB Assessment (Version 3.0)*; European Chemicals Agency: Helsinki, Finland, 2017.

(12) Wania, F.; Lei, Y. D.; Baskaran, S.; Sangion, A. Identifying organic chemicals not subject to bioaccumulation in air-breathing organisms using predicted partitioning and biotransformation properties. *Integr. Environ. Assess. Manag.* **2021**, DOI: 10.1002/ieam.4555.

(13) Zhang, Z.; Wang, S.; Li, L. The role of chemical properties in human exposure to environmental chemicals. *Environ Sci Process Impacts* **2021**, *23*, 1839−1862.

(14) Zhang, X.; Brown, T. N.; Wania, F.; Heimstad, E. S.; Goss, K. U. Assessment of chemical screening outcomes based on different partitioning property estimation methods. *Environ Int* **2010**, *36* (6), 514−520.

(15) Odziomek, K.; Gajewicz, A.; Haranczyk, M.; Puzyn, T. Reliability of environmental fate modeling results for POPs based on various methods of determining the air/water partition coefficient (log K_{AW}). *Atmos. Environ.* **2013**, *73*, 177−184.

(16) Organisation for Economic Co-operation and Development (OECD). *Guidance Document on the Use of Multimedia Models for Estimating Overall Environmental Persistance and Long-Range Transport*; Organisation for Economic Co-operation and Development: Paris, 2004.

(17) Organisation for Economic Cooperation and Development (OECD). *Report of the OECD/UNEP Workshop on the Use of Multimedia Models for Estimating Overall Environmental Persistence and Long Range Transport in the Context of PBTS/POPS Assessment*, 2002; https://www.oecd-ilibrary.org/environment/report-of-the-oecd-unep-workshop-on-the-use-of-multimedia-models-for-estimating-overall-environmental-persistence-and-long-range-transport-in-the-context-of-pbts-pops-assessment_9789264078505-en (accessed 2022-07-12).

(18) Buser, A. M.; MacLeod, M.; Scheringer, M.; Mackay, D.; Bonnell, M.; Russell, M. H.; DePinto, J. V.; Hungerbühler, K. Good modeling practice guidelines for applying multimedia models in chemical assessments. *Integr. Environ. Assess. Manag.* **2012**, *8* (4), 703−708.

(19) Pontolillo, J.; Eganhouse, R. P. *The Search for Reliable Aqueous Solubility (Sw) and Octanol-water Partition Coefficient (Kow) Data for Hydrophobic Organic Compounds: DDT and DDE as A Case Study*; U.S. Department of the Interior, U.S. Geological Survey, 2001.

(20) Aurisano, N.; Fantke, P. Semi-automated harmonization and selection of chemical data for risk and impact assessment. *Chemosphere* **2022**, *302*, 134886.

(21) Organisation for Economic Co-operation and Development (OECD). *OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models*; Organisation for Economic Co-operation and Development: Paris, 2004.

(22) Klamt, A.; Eckert, F.; Arlt, W. COSMO-RS: an alternative to simulation for calculating thermodynamic properties of liquid mixtures. *Annu. Rev. Chem. Biomol. Eng.* **2010**, *1*, 101−122.

(23) Zhu, H.; Tropsha, A.; Fourches, D.; Varnek, A.; Papa, E.; Gramatica, P.; Öberg, T.; Dao, P.; Cherkasov, A.; Tetko, I. V.

Combinatorial QSAR modeling of chemical toxicants tested against Tetrahymena pyriformis. *J. Chem. Inf. Model.* **2008**, *48* (4), 766−784.
(24) Boethling, R. S.; Howard, P. H.; Meylan, W. M. Finding and estimating chemical property data for environmental assessment. *Environ. Toxicol. Chem.* **2004**, *23* (10), 2290−2308.
(25) Mackay, D.; Celsie, A. K.; Parnis, J. M. The evolution and future of environmental partition coefficients. *Environ. Rev.* **2016**, *24* (1), 101−113.
(26) Schwarzenbach, R. P.; Gschwend, P. M.; Imboden, D. M. *Environmental Organic Chemistry*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, 2017.
(27) International Union of Pure and Applied Chemistry (IUPAC). *Compendium of Chemical Terminology (the "Gold Book")*, 2nd ed; compiled by McNaught, A. D.; Wilkinson, A.; Blackwell Scientific Publications: Oxford, 1997.
(28) Cole, J. G.; Mackay, D. Correlating environmental partitioning properties of organic compounds: The three solubility approach. *Environ. Toxicol. Chem.* **2000**, *19* (2), 265−270.
(29) Mackay, D. *Multimedia Environmental Models: The Fugacity Approach*; CRC Press: Boca Raton, FL, 2001.
(30) Baskaran, S.; Lei, Y. D.; Wania, F. A database of experimentally derived and estimated octanol−air partition ratios (K_{OA}). *J. Phys. Chem. Ref. Data* **2021**, *50* (4), 043101.
(31) Sander, R.; Acree, W. E.; De Visscher, A.; Schwartz, S. E.; Wallington, T. J. Henry's law constants (IUPAC Recommendations 2021). *Pure Appl. Chem.* **2022**, *94* (1), 71−85.
(32) Matthies, M.; Beulke, S. Considerations of temperature in the context of the persistence classification in the EU. *Environ. Sci. Eur.* **2017**, *29*, 15.
(33) Goss, K.-U.; Eisenreich, S. J. Adsorption of VOCs from the gas phase to different minerals and a mineral mixture. *Environ. Sci. Technol.* **1996**, *30* (7), 2135−2142.
(34) Sangster, J. M. *Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry*; John Wiley & Sons, 1997; Vol. *1*.
(35) Leo, A.; Hansch, C.; Elkins, D. Partition coefficients and their uses. *Chem. Rev.* **1971**, *71* (6), 525−616.
(36) Mackay, D.; Shiu, W. Y. A critical review of Henry's law constants for chemicals of environmental interest. *J. Phys. Chem. Ref. Data* **1981**, *10* (4), 1175−1199.
(37) Mackay, D.; Shiu, W.-Y.; Lee, S. C. *Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*; CRC Press: Boca Raton, FL, 2006.
(38) Yalkowsky, S. H.; Banerjee, S. *Aqueous Solubility: Methods of Estimation for Organic Compounds*; Marcel Dekker: New York, 1992.
(39) Yalkowsky, S. H.; He, Y.; Jain, P. *Handbook of Aqueous Solubility Data*, 2nd ed.; CRC Press: Boca Raton, FL, 2016.
(40) Wilhoit, R. C.; Zwolinski, B. J. *Handbook of Vapor Pressures and Heats of Vaporization of Hydrocarbons and Related Compounds*; Thermodynamics Research Center, Department of Chemistry, Texas A&M University: College Station, TX, 1971.
(41) Daubert, T. E.; Danner, R. P. *Physical and Thermodynamic Properties of Pure Chemicals*; Taylor & Francis: Washington, D.C., 1989.
(42) Sander, R. Compilation of Henry's law constants (version 4.0) for water as solvent. *Atmos. Chem. Phys.* **2015**, *15* (8), 4399−4981.
(43) Kollig, H. P. Criteria for evaluating the reliability of literature data on environmental process constants. *Toxicol. Environ. Chem.* **1988**, *17* (4), 287−311.
(44) Mackay, D.; Myland, J. C.; Oldham, K. B.; Parnis, J. M. Pitfalls in the application of statistics to chemical data: The determination of the partition ratio K_{OW} as a case in point. *J. Math. Chem.* **2018**, *56* (5), 1407−1427.
(45) Parnis, J. M.; Taskovic, T.; Celsie, A. K. D.; Mackay, D. Indoor dust/air partitioning: Evidence for kinetic delay in equilibration for low-volatility SVOCs. *Environ. Sci. Technol.* **2020**, *54*, 6723.
(46) Men, Y.; Achermann, S.; Helbling, D. E.; Johnson, D. R.; Fenner, K. Relative contribution of ammonia oxidizing bacteria and other members of nitrifying activated sludge communities to micropollutant biotransformation. *Water Res.* **2017**, *109*, 217−226.
(47) Wania, F.; Lei, Y. D.; Harner, T. Estimating octanol−air partition coefficients of nonpolar semivolatile organic compounds from gas chromatographic retention times. *Anal. Chem.* **2002**, *74* (14), 3476−3483.
(48) Rodgers, T. F.; Okeme, J. O.; Bidleman, T. F. Comment on "A database of experimentally derived and estimated octanol−air partition ratios (K_{OA})". *J. Phys. Chem. Ref. Data* **2022**, *51* (2), 026101.
(49) Baskaran, S.; Lei, Y. D.; Wania, F. Response to comment on "A database of experimentally derived and estimated octanol−air partition ratios (K_{OA})". *J. Phys. Chem. Ref. Data* **2022**, *51* (2), 026102.
(50) Klein, W.; Kördel, W.; Weiss, M.; Poremski, H. Updating of the OECD test guideline 107 "partition coefficient n-octanol/water": OECD laboratory intercomparison test on the HPLC method. *Chemosphere* **1988**, *17* (2), 361−386.
(51) Shunthirasingham, C.; Lei, Y. D.; Wania, F. Evidence of bias in air−water Henry's Law Constants for semivolatile organic compounds measured by inert gas stripping. *Environ. Sci. Technol.* **2007**, *41* (11), 3807−3814.
(52) Lei, Y. D.; Shunthirasingham, C.; Wania, F. Comparison of headspace and gas-stripping techniques for measuring the air-water partitioning of normal alkanols (C4 to C10): Effect of temperature, chain length, and adsorption to the water surface. *J. Chem. Eng. Data* **2007**, *52* (1), 168−179.
(53) Shunthirasingham, C.; Cao, X.; Lei, Y. D.; Wania, F. Large bubbles reduce the surface sorption artifact of the inert gas stripping method. *J. Chem. Eng. Data* **2013**, *58* (3), 792−797.
(54) Fenner, K.; Canonica, S.; Wackett, L. P.; Elsner, M. Evaluating pesticide degradation in the environment: blind spots and emerging opportunities. *Science* **2013**, *341* (6147), 752−758.
(55) Gulde, R.; Anliker, S.; Kohler, H. E.; Fenner, K. Ion trapping of amines in protozoa: A novel removal mechanism for micropollutants in activated sludge. *Environ. Sci. Technol.* **2018**, *52* (1), 52−60.
(56) Luo, Y.; Guo, W.; Ngo, H. H.; Nghiem, L. D.; Hai, F. I.; Zhang, J.; Liang, S.; Wang, X. C. A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Sci. Total Environ.* **2014**, *473*−*474*, 619−641.
(57) Li, B.; Zhang, T. Biodegradation and adsorption of antibiotics in the activated sludge process. *Environ. Sci. Technol.* **2010**, *44* (9), 3468−3473.
(58) Lindqvist, R.; Enfield, C. G. Biosorption of dichlorodiphenyltrichloroethane and hexachlorobenzene in groundwater and its implications for facilitated transport. *Appl. Environ. Microbiol.* **1992**, *58* (7), 2211−2218.
(59) Falas, P.; Jewell, K. S.; Hermes, N.; Wick, A.; Ternes, T. A.; Joss, A.; Nielsen, J. L. Transformation, CO_2 formation and uptake of four organic micropollutants by carrier-attached microorganisms. *Water Res.* **2018**, *141*, 405−416.
(60) Organisation for Economic Co-operation and Development (OECD). *Test No. 107: Partition Coefficient (n-octanol/water): Shake Flask Method*; Organisation for Economic Co-operation and Development: Paris, 2006.
(61) Organisation for Economic Co-operation and Development (OECD). *Test No. 117: Partition Coefficient (n-octanol/water): HPLC Method*; Organisation for Economic Co-operation and Development: Paris, 2004.
(62) Organisation for Economic Co-operation and Development (OECD). *Test No. 123: Partition Coefficient (1-Octanol/Water): Slow-Stirring Method*; Organisation for Economic Co-operation and Development: Paris, 2006.
(63) Lei, Y. D.; Baskaran, S.; Wania, F. Measuring the octan-1-ol air partition coefficient of volatile organic chemicals with the variable phase ratio headspace technique. *J. Chem. Eng. Data* **2019**, *64* (11), 4793−4800.
(64) Ha, Y.; Kwon, J.-H. Determination of 1-octanol-air partition coefficient using gaseous diffusion in the air boundary layer. *Environ. Sci. Technol.* **2010**, *44* (8), 3041−3046.

(65) Birch, H.; Redman, A. D.; Letinski, D. J.; Lyon, D. Y.; Mayer, P. Determining the water solubility of difficult-to-test substances: a tutorial review. *Anal. Chim. Acta* **2019**, *1086*, 16−28.
(66) Organisation for Economic Co-operation and Development (OECD). *Test No. 104: Vapour Pressure*; Organisation for Economic Co-operation and Development: Paris, 2006.
(67) Howard, P. H.; Banerjee, S.; Robillard, K. H. Measurement of water solubilities, octanol/water partition coefficients and vapor pressures of commercial phthalate esters. *Environ. Toxicol. Chem.* **1985**, *4* (5), 653−661.
(68) Harnisch, M.; Möckel, H. J.; Schulze, G. Relationship between log Pow, shake-flask values and capacity factors derived from reversed-phase high-performance liquid chromatography for *n*-alkylbenzenes and some OECD reference substances. *J. Chromatogr. A* **1983**, *282*, 315−332.
(69) Brooke, D.; Nielsen, I.; de Bruijn, J.; Hermens, J. An interlaboratory evaluation of the stir-flask method for the determination of octanol-water partition coefficients (Log Pow). *Chemosphere* **1990**, *21* (1−2), 119−133.
(70) De Bruijn, J.; Busser, F.; Seinen, W.; Hermens, J. Determination of octanol/water partition coefficients for hydrophobic organic chemicals with the "slow-stirring" method. *Environ. Toxicol. Chem.* **1989**, *8* (6), 499−512.
(71) Ellington, J. J.; Floyd, T. L. *Octanol/Water Partition Coefficients for Eight Phthalate Esters. Technical Report EPA/600/S-96/006*; U.S. Environmental Protection Agency, National Exposure Research Laboratory: Cincinnati, OH, 1996.
(72) Pegoraro, C. N.; Chiappero, M. S.; Montejano, H. A. Measurements of octanol−air partition coefficients, vapor pressures and vaporization enthalpies of the (E) and (Z) isomers of the 2-ethylhexyl 4-methoxycinnamate as parameters of environmental impact assessment. *Chemosphere* **2015**, *138*, 546−552.
(73) Hinckley, D. A.; Bidleman, T. F.; Foreman, W. T.; Tuschall, J. R. Determination of vapor pressures for nonpolar and semipolar organic compounds from gas chromatograhic retention data. *J. Chem. Eng. Data* **1990**, *35* (3), 232−237.
(74) Small, P.; Small, K.; Cowley, P. The vapour pressures of some high boiling esters. *Transactions of the Faraday Society* **1948**, *44*, 810−816.
(75) Perry, E.; Weber, W. Vapor pressures of phlegmatic liquids. II. High molecular weight esters and silicone oils. *J. Am. Chem. Soc.* **1949**, *71* (11), 3726−3730.
(76) Werner, A. Vapor pressures of phthalate esters. *Industrial & Engineering Chemistry* **1952**, *44* (11), 2736−2740.
(77) Gobble, C.; Chickos, J.; Verevkin, S. P. Vapor pressures and vaporization enthalpies of a series of dialkyl phthalates by correlation gas chromatography. *J. Chem. Eng. Data* **2014**, *59* (4), 1353−1365.
(78) Liang, Y.; Xu, Y. Improved method for measuring and characterizing phthalate emissions from building materials and its application to exposure assessment. *Environ. Sci. Technol.* **2014**, *48* (8), 4475−4484.
(79) Hollifield, H. C. Rapid nephelometric estimate of water solubility of highly insoluble organic chemicals of environmental interest. *Bull. Environ. Contam. Toxicol.* **1979**, *23* (1), 579−586.
(80) Leyder, F.; Boulanger, P. Ultraviolet absorption, aqueous solubility, and octanol-water partition for several phthalates. *Bull. Environ. Contam. Toxicol.* **1983**, *30* (1), 152−157.
(81) Wolfe, N.; Steen, W.; Burns, L. Phthalate ester hydrolysis: linear free energy relationships. *Chemosphere* **1980**, *9* (7−8), 403−408.
(82) Defoe, D. L.; Holcombe, G. W.; Hammermeister, D. E.; Biesinger, K. E. Solubility and toxicity of eight phthalate esters to four aquatic organisms. *Environ. Toxicol. Chem.* **1990**, *9* (5), 623−636.
(83) Thomsen, M.; Carlsen, L.; Hvidt, S. Solubilities and surface activities of phthalates investigated by surface tension measurements. *Environ. Toxicol. Chem.* **2001**, *20* (1), 127−132.
(84) Letinski, D. J.; Connelly, M. J., Jr; Peterson, D. R.; Parkerton, T. F. Slow-stir water solubility measurements of selected alcohols and diesters. *Chemosphere* **2002**, *48* (3), 257−265.
(85) Sugatt, R. H.; O'Grady, D. P.; Banerjee, S.; Howard, P. H.; Gledhill, W. E. Shake flask biodegradation of 14 commercial phthalate esters. *Appl. Environ. Microbiol.* **1984**, *47* (4), 601−606.
(86) Cartwright, C. D.; Thompson, I. P.; Burns, R. G. Degradation and impact of phthalate plasticizers on soil microbial communities. *Environ. Toxicol. Chem.* **2000**, *19* (5), 1253−1261.
(87) Juneson, C.; Ward, O. P.; Singh, A. Biodegradation of bis(2-ethylhexyl)phthalate in a soil slurry-sequencing batch reactor. *Process Biochem.* **2001**, *37* (3), 305−313.
(88) Xu, G.; Li, F.; Wang, Q. Occurrence and degradation characteristics of dibutyl phthalate (DBP) and di-(2-ethylhexyl) phthalate (DEHP) in typical agricultural soils of China. *Sc. Total Environ.* **2008**, *393* (2), 333−340.
(89) Yuwatini, E.; Hata, N.; Taguchi, S. Behavior of di(2-ethylhexyl) phthalate discharged from domestic waste water into aquatic environment. *J. Environ. Monit.* **2006**, *8* (1), 191−196.
(90) Yuan, S. Y.; Liu, C.; Liao, C. S.; Chang, B. V. Occurrence and microbial degradation of phthalate esters in Taiwan river sediments. *Chemosphere* **2002**, *49* (10), 1295−1299.
(91) Yuan, S. Y.; Chang, J. S.; Yen, J. H.; Chang, B.-V. Biodegradation of phenanthrene in river sediment. *Chemosphere* **2001**, *43* (3), 273−278.
(92) Li, N.; Wania, F.; Lei, Y. D.; Daly, G. L. A comprehensive and critical compilation, evaluation, and selection of physical−chemical property data for selected polychlorinated biphenyls. *J. Phys. Chem. Ref. Data* **2003**, *32* (4), 1545−1590.
(93) Åberg, A.; MacLeod, M.; Wiberg, K. Physical-chemical property data for dibenzo-p-dioxin (DD), dibenzofuran (DF), and chlorinated DD/Fs: A critical review and recommended values. *J. Phys. Chem. Ref. Data* **2008**, *37* (4), 1997−2008.
(94) Beyer, A.; Wania, F.; Gouin, T.; Mackay, D.; Matthies, M. Selecting internally consistent physicochemical properties of organic compounds. *Environ. Toxicol. Chem.* **2002**, *21* (5), 941−953.
(95) Shen, L.; Wania, F. Compilation, evaluation, and selection of physical− chemical property data for organochlorine pesticides. *J. Chem. Eng. Data* **2005**, *50* (3), 742−768.
(96) Schenker, U.; Soltermann, F.; Scheringer, M.; Hungerbühler, K. Modeling the environmental fate of polybrominated diphenyl ethers (PBDEs): The importance of photolysis for the formation of lighter PBDEs. *Environ. Sci. Technol.* **2008**, *42* (24), 9244−9249.
(97) Arnot, J.; McCarty, L.; Armitage, J.; Toose-Reid, L.; Wania, F.; Cousins, I. *An Evaluation of Hexabromocyclododecane (HBCD) for Persistent Organic Pollutant (POP) Properties and the Potential for Adverse Effects in the Environment. A report submitted to European Brominated Flame Retardant Industry Panel (EBFRIP)*, 2009; https://unece.org/fileadmin/DAM/env/documents/2009/EB/wg5/wgsr45/Informal%20docs/An%20evaluation%20of%20hexabromocyclododecane_Final%20report.pdf (accessed 2022-07-12).
(98) Buser, A. M.; Schenker, S.; Scheringer, M.; Hungerbühler, K. Comparing the performance of computational estimation methods for physicochemical properties of dimethylsiloxanes and selected siloxanols. *J. Chem. Eng. Data* **2013**, *58* (11), 3170−3178.
(99) Ma, Y.-G.; Lei, Y. D.; Xiao, H.; Wania, F.; Wang, W.-H. Critical review and recommended values for the physical-chemical property data of 15 polycyclic aromatic hydrocarbons at 25 °C. *J. Chem. Eng. Data* **2010**, *55* (2), 819−825.
(100) Li, L.; Liu, J.; Hu, J. Global inventory, long-range transport and environmental distribution of dicofol. *Environ. Sci. Technol.* **2015**, *49* (1), 212−222.
(101) European Chemicals Agency. *Guidance on Information Requirements and Chemical Safety Assessment. Chapter R.7b: Endpoint Specific Guidance (Version 4.0)*; European Chemicals Agency: Helsinki, Finland, 2017.
(102) Panel on Plant Protection Products and their Residues, Opinion on a rrequest from EFSA related to the default Q10 value used to describe the temperature effect on transformation rates of pesticides in soil. *The EFSA Journal* **2007**, *622*, 1−32.

(103) Anderson, P. N.; Hites, R. A. OH radical reactions: The major removal pathway for polychlorinated biphenyls from the atmosphere. *Environ. Sci. Technol.* **1996**, *30* (5), 1756−1763.
(104) Fenner, K.; Scheringer, M.; MacLeod, M.; Matthies, M.; McKone, T.; Stroebe, M.; Beyer, A.; Bonnell, M.; Le Gall, A. C.; Klasmeier, J.; et al. Comparing estimates of persistence and long-range transport potential among multimedia models. *Environ. Sci. Technol.* **2005**, *39* (7), 1932−1942.
(105) Boethling, R. S.; Howard, P. H.; Beauman, J. A.; Larosch, M. E. Factors for intermedia extrapolation in biodegradability assessment. *Chemosphere* **1995**, *30* (4), 741−752.
(106) Aronson, D.; Boethling, R.; Howard, P.; Stiteler, W. Estimating biodegradation half-lives for use in chemical screening. *Chemosphere* **2006**, *63* (11), 1953−1960.
(107) Fenner, K.; Screpanti, C.; Renold, P.; Rouchdi, M.; Vogler, B.; Rich, S. Comparison of small molecule biotransformation half-lives between activated sludge and soil: Opportunities for read-across? *Environ. Sci. Technol.* **2020**, *54* (6), 3148−3158.
(108) MacLeod, M.; Scheringer, M.; Hungerbühler, K. Estimating enthalpy of vaporization from vapor pressure using Trouton's Rule. *Environ. Sci. Technol.* **2007**, *41* (8), 2827−2832.
(109) Goss, K.-U.; Schwarzenbach, R. P. Empirical prediction of heats of vaporization and heats of adsorption of organic compounds. *Environ. Sci. Technol.* **1999**, *33* (19), 3390−3393.
(110) Karickhoff, S. W.; Brown, D. S.; Scott, T. A. Sorption of hydrophobic pollutants on natural sediments. *Water Res.* **1979**, *13* (3), 241−248.
(111) Karickhoff, S. W. Semi-empirical estimation of sorption of hydrophobic pollutants on natural sediments and soils. *Chemosphere* **1981**, *10* (8), 833−846.
(112) Seth, R.; Mackay, D.; Muncke, J. Estimating the organic carbon partition coefficient and its variability for hydrophobic chemicals. *Environ. Sci. Technol.* **1999**, *33* (14), 2390−2394.
(113) Gerstl, Z. Estimation of organic chemical sorption by soils. *J. Contam. Hydrol.* **1990**, *6* (4), 357−375.
(114) Goss, K.-U.; Schwarzenbach, R. P. Linear free energy relationships used to evaluate equilibrium partitioning of organic compounds. *Environ. Sci. Technol.* **2001**, *35* (1), 1−9.
(115) Droge, S. T.; Goss, K.-U. Sorption of organic cations to phyllosilicate clay minerals: CEC-normalization, salt dependency, and the role of electrostatic and hydrophobic effects. *Environ. Sci. Technol.* **2013**, *47* (24), 14224−14232.
(116) Droge, S. T.; Goss, K.-U. Development and evaluation of a new sorption model for organic cations in soil: contributions from organic matter and clay minerals. *Environ. Sci. Technol.* **2013**, *47* (24), 14233−14241.
(117) Karickhoff, S. W. Organic pollutant sorption in aquatic systems. *J. Hydraul. Eng.* **1984**, *110* (6), 707−735.
(118) Xiao, H.; Wania, F. Is vapor pressure or the octanol−air partition coefficient a better descriptor of the partitioning between gas phase and organic matter? *Atmos. Environ.* **2003**, *37* (20), 2867−2878.
(119) MacLeod, M.; Scheringer, M.; Götz, C.; Hungerbühler, K.; Davidson, C.; Holsen, T. Deposition from the atmosphere to water and soils with aerosol particles and precipitation. In *Handbook of Chemical Mass Transport in the Environment*; Thibodeaux, L. J., Mackay, D., Eds.; CRC Press: Boca Raton, FL, 2011.
(120) Ran, Y.; He, Y.; Yang, G.; Johnson, J. L. H.; Yalkowsky, S. H. Estimation of aqueous solubility of organic compounds by using the general solubility equation. *Chemosphere* **2002**, *48* (5), 487−509.
(121) Endo, S.; Escher, B. I.; Goss, K.-U. Capacities of membrane lipids to accumulate neutral organic chemicals. *Environ. Sci. Technol.* **2011**, *45* (14), 5912−5921.
(122) Endo, S.; Bauerfeind, J.; Goss, K.-U. Partitioning of neutral organic compounds to structural proteins. *Environ. Sci. Technol.* **2012**, *46* (22), 12697−12703.
(123) Reppas-Chrysovitsinos, E.; Sobek, A.; MacLeod, M. Screening-level models to estimate partition ratios of organic chemicals between polymeric materials, air and water. *Environ. Sci. Processes Impacts* **2016**, *18* (6), 667−676.
(124) Harner, T.; Bidleman, T. F. Octanol-air partition coefficient for describing particle/gas partitioning of aromatic compounds in urban air. *Environ. Sci. Technol.* **1998**, *32* (10), 1494−1502.
(125) Saini, A.; Okeme, J.; Mark Parnis, J.; McQueen, R.; Diamond, M. From air to clothing: characterizing the accumulation of semi-volatile organic compounds to fabrics in indoor environments. *Indoor air* **2017**, *27* (3), 631−641.
(126) Saini, A.; Rauert, C.; Simpson, M. J.; Harrad, S.; Diamond, M. L. Characterizing the sorption of polybrominated diphenyl ethers (PBDEs) to cotton and polyester fabrics under controlled conditions. *Sci. Total Environ.* **2016**, *563*, 99−107.
(127) Csiszar, S. A.; Diamond, M. L.; Thibodeaux, L. J. Modeling urban films using a dynamic multimedia fugacity model. *Chemosphere* **2012**, *87* (9), 1024−1031.
(128) Endo, S.; Goss, K.-U. Applications of polyparameter linear free energy relationships in environmental chemistry. *Environ. Sci. Technol.* **2014**, *48* (21), 12477−12491.
(129) Breivik, K.; Wania, F. Expanding the applicability of multimedia fate models to polar organic chemicals. *Environ. Sci. Technol.* **2003**, *37* (21), 4934−4943.
(130) Endo, S.; Goss, K.-U. Predicting partition coefficients of polyfluorinated and organosilicon compounds using polyparameter linear free energy relationships (PP-LFERs). *Environ. Sci. Technol.* **2014**, *48* (5), 2776−2784.
(131) Baskaran, S.; Lei, Y. D.; Wania, F. Reliable prediction of the octanol−air partition ratio. *Environ. Toxicol. Chem.* **2021**, *40* (11), 3166−3180.
(132) Goss, K.-U. Prediction of the temperature dependency of Henry's law constant using poly-parameter linear free energy relationships. *Chemosphere* **2006**, *64* (8), 1369−1374.
(133) Mintz, C.; Burton, K.; Ladlie, T.; Clark, M.; Acree, W. E.; Abraham, M. H. Enthalpy of solvation correlations for gaseous solutes dissolved in dibutyl ether and ethyl acetate. *Thermochim. Acta* **2008**, *470* (1), 67−76.
(134) Bronner, G.; Goss, K.-U. Predicting sorption of pesticides and other multifunctional organic chemicals to soil organic carbon. *Environ. Sci. Technol.* **2011**, *45* (4), 1313−1319.
(135) Geisler, A.; Endo, S.; Goss, K.-U. Partitioning of organic chemicals to storage lipids: elucidating the dependence on fatty acid composition and temperature. *Environ. Sci. Technol.* **2012**, *46* (17), 9519−9524.
(136) U.S. Environmental Protection Agency. *Estimation Programs Interface (EPI) Suite for Microsoft® Windows, v 4.1*; U.S. Environmental Protection Agency: Washington, D.C., 2012.
(137) Mansouri, K.; Grulke, C. M.; Judson, R. S.; Williams, A. J. OPERA models for predicting physicochemical properties and environmental fate endpoints. *J. Cheminformatics* **2018**, *10* (1), 10.
(138) Gramatica, P.; Cassani, S.; Chirico, N. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *J. Comput. Chem.* **2014**, *35* (13), 1036−1044.
(139) Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem.* **2013**, *34* (24), 2121−2132.
(140) Brown, T. N.; Armitage, J. M.; Arnot, J. A. Application of an Iterative Fragment Selection (IFS) method to estimate entropies of fusion and melting points of organic chemicals. *Mol. Inform.* **2019**, *38* (8−9), 1800160.
(141) Brown, T. N.; Arnot, J. A.; Wania, F. Iterative fragment selection: A group contribution approach to predicting fish biotransformation half-lives. *Environ. Sci. Technol.* **2012**, *46* (15), 8253−8260.
(142) Arnot, J. A.; Brown, T. N.; Wania, F. Estimating screening-level organic chemical half-lives in humans. *Environ. Sci. Technol.* **2014**, *48* (1), 723−730.
(143) Brown, T. N. QSPRs for predicting equilibrium partitioning in solvent−air systems from the chemical structures of solutes and solvents. *J. Solution Chem.* **2022**, DOI: 10.1007/s10953-022-01162-2.

(144) Meylan, W. M.; Howard, P. H. Atom/fragment contribution method for estimating octanol−water partition coefficients. *J. Pharm. Sci.* **1995**, *84* (1), 83−92.
(145) Organisation for Economic Co-operation and Development (OECD). *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*; Organisation for Economic Co-operation and Development: Paris, 2007.
(146) Sarfraz Iqbal, M.; Golsteijn, L.; Öberg, T.; Sahlin, U.; Papa, E.; Kovarich, S.; Huijbregts, M. A. J. Understanding quantitative structure-property relationships uncertainty in environmental fate modeling. *Environ. Toxicol. Chem.* **2013**, *32* (5), 1069−1076.
(147) Hanser, T.; Barber, C.; Marchaland, J.; Werner, S. Applicability domain: towards a more formal definition. *SAR and QSAR in Environ. Res.* **2016**, *27* (11), 865−881.
(148) Dearden, J.; Cronin, M. T.; Kaiser, K. L. How not to develop a quantitative structure−activity or structure−property relationship (QSAR/QSPR). *SAR QSAR Environ. Res.* **2009**, *20* (3−4), 241−266.
(149) Puzyn, T. On the replacement of empirical parameters in multimedia mass balance models with QSPR data. *J. Hazard. Mat.* **2011**, *192* (3), 970−977.
(150) Rodgers, T. F.; Okeme, J. O.; Parnis, J. M.; Girdhari, K.; Bidleman, T. F.; Wan, Y.; Jantunen, L. M.; Diamond, M. L. Novel Bayesian method to derive final adjusted values of physicochemical properties: Application to 74 compounds. *Environ. Sci. Technol.* **2021**, *55* (18), 12302−12316.
(151) Lampic, A.; Parnis, J. M. Property estimation of per- and polyfluoroalkyl substances: A comparative assessment of estimation methods. *Environ. Toxicol. Chem.* **2020**, *39* (4), 775−786.
(152) Arnot, J. A.; Gouin, T.; Mackay, D. *Practical Methods for Estimating Environmental Biodegradation Rates. Report for Environment Canada (CEMN2005003)*; Canadian Environmental Modelling Network, Trent University: Peterborough, ON, Canada, 2005.
(153) Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inf.* **2010**, *29* (6−7), 476−488.
(154) Tebes-Stevens, C.; Patel, J. M.; Koopmans, M.; Olmstead, J.; Hilal, S. H.; Pope, N.; Weber, E. J.; Wolfe, K. Demonstration of a consensus approach for the calculation of physicochemical properties required for environmental fate assessments. *Chemosphere* **2018**, *194*, 94−106.
(155) Gramatica, P.; Pilutti, P.; Papa, E. Validated QSAR prediction of OH tropospheric degradation of VOCs: splitting into training− test sets and consensus modeling. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (5), 1794−1802.
(156) Endo, S. Applicability domain of polyparameter linear free energy relationship models evaluated by leverage and prediction interval calculation. *Environ. Sci. Technol.* **2022**, *56* (9), 5572−5579.
(157) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31−36.
(158) Mansouri, K.; Grulke, C.; Richard, A.; Judson, R.; Williams, A. An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling. *SAR QSAR Environ. Res.* **2016**, *27* (11), 911−937.
(159) Williams, A. J.; Ekins, S.; Tkachenko, V. Towards a gold standard: regarding quality in public domain chemistry databases and approaches to improving the situation. *Drug discovery today* **2012**, *17* (13−14), 685−701.
(160) Williams, A. J.; Ekins, S. A quality alert and call for improved curation of public chemistry databases. *Drug Discovery Today* **2011**, *16* (17−18), 747−750.
(161) Young, D.; Martin, T.; Venkatapathy, R.; Harten, P. Are the chemical structures in your QSAR correct? *QSAR Comb. Sci.* **2008**, *27* (11−12), 1337−1345.
(162) Waldman, M.; Fraczkiewicz, R.; Clark, R. D. Tales from the war on error: the art and science of curating QSAR data. *J. Comput. Aided Mol. Des.* **2015**, *29* (9), 897−910.
(163) Gadaleta, D.; Lombardo, A.; Toma, C.; Benfenati, E. A new semi-automated workflow for chemical data retrieval and quality checking for modeling applications. *J. Cheminform.* **2018**, *10* (1), 60.
(164) Martinez, M. N.; Bartholomew, M. J. What does it "mean"? A review of interpreting and calculating different types of means and standard deviations. *Pharmaceutics* **2017**, *9* (2), 14.
(165) Hofer, T.; Myhre, O.; Peltola-Thies, J.; Hirmann, D. Analysis of elimination half-lives in MamTKDB 1.0 related to bioaccumulation: Requirement of repeated administration and blood plasma values underrepresent tissues. *Environ. Int.* **2021**, *155*, 106592.
(166) Xiao, H.; Li, N.; Wania, F. Compilation, evaluation, and selection of physical-chemical property data for α-, β-, and γ-hexachlorocyclohexane. *J. Chem. Eng. Data* **2004**, *49* (2), 173−185.
(167) Wania, F.; Dugani, C. B. Assessing the long-range transport potential of polybrominated diphenyl ethers: A comparison of four multimedia models. *Environ. Toxicol. Chem.* **2003**, *22* (6), 1252−1261.
(168) Yue, C.; Li, L. Y. Filling the gap: Estimating physicochemical properties of the full array of polybrominated diphenyl ethers (PBDEs). *Environ. Pollut.* **2013**, *180*, 312−323.

https://doi.org/10.1021/acsenvironau.2c00010
ACS Environ. Au 2022, 2, 376−395