Identifying and designing chemicals with minimal acute aquatic toxicity

Jakub Kostal[a], Adelina Voutchkova-Kostal[b], Paul T. Anastas[c], and Julie Beth Zimmerman[c,d,1]

[b]Department of Chemistry, George Washington University, Washington, DC 20052; [a]Sustainability A to Z, LLC, Guilford, CT 06437; and [c]School of Forestry and Environmental Studies and [d]Department of Chemical and Environmental Engineering, Yale University, New Haven, CT 06511

Industrial ecology has revolutionized our understanding of material stocks and flows in our economy and society. For this important discipline to have even deeper impact, we must understand the inherent nature of these materials in terms of human health and the environment. This paper focuses on methods to design synthetic chemicals to reduce their intrinsic ability to cause adverse consequence to the biosphere. Advances in the fields of computational chemistry and molecular toxicology in recent decades allow the development of predictive models that inform the design of molecules with reduced potential to be toxic to humans or the environment. The approach presented herein builds on the important work in quantitative structure–activity relationships by linking toxicological and chemical mechanistic insights to the identification of critical physical–chemical properties needed to be modified. This in silico approach yields design guidelines using boundary values for physiochemical properties. Acute aquatic toxicity serves as a model endpoint in this study. Defining value ranges for properties related to bioavailability and reactivity eliminates 99% of the chemicals in the highest concern for acute aquatic toxicity category. This approach and its future implementations are expected to yield very powerful tools for life cycle assessment practitioners and molecular designers that allow rapid assessment of multiple environmental and human health endpoints and inform modifications to minimize hazard.

green chemistry | safer chemicals | rational design | toxicity prediction

Industrial ecology and green chemistry are two rigorous scientific disciplines with global scientific communities that empower sustainability science. Sustainability science is the science, technology, and innovation in support of sustainable development—meeting human needs and reducing hunger and poverty while maintaining the life support systems of the planet (1, 2). With a systems view, industrial ecology investigates material and energy flows of coupled human–natural systems and has made significant strides in assessing the impacts of these flows on the environment and human health (3–8). The need for more sustainable products and processes has triggered (further) development of a large number of environmental assessment tools (9), including substance flow analysis (10), chemical/product risk assessment (11), life cycle assessment (LCA) (12–14), and a variety of screening tools (15–19). The knowledge generated by these investigations and assessments provides key information about the chemicals, materials and processes with the most significant adverse impacts throughout the life cycle. We need to understand the inherent nature of these materials to not only quantify their impact on human health and the environment but also to facilitate the design of a more sustainable materials basis of our society. Analogous to the industrial ecology assessment tools, several National Academies of Science reports have identified the need for new green chemistry design tools (20–22), and specifically, tools focused on molecular design for reduced toxicity (23). Although the majority of commercial chemicals are not intended to be biologically active, many have reported unintended biological activity that leads to a wide range of human health

and ecotoxicological impacts (11, 24). It has become increasingly evident that there are significant concerns about the adverse human health and ecosystem impacts resulting from chemical exposure and the challenge associated with predicting and modeling such endpoints (25). Several tools have emerged in this space with a consensus-building effort around USETox (26–28). Many of these tools rely on the inherent nature of the chemicals being assessed, such as the octanol-water partition coefficient, as well as circumstantial information related to fate, transport, and exposure.

Extensive human health and ecotoxicological testing of all new chemicals to determine inherent toxicity characteristics is not feasible due to the number of new substances introduced daily, the time it takes to conduct reviews, and the prohibitive economic and social costs of testing, particularly in vivo (29). These concerns could be mitigated by addressing the significant challenge of designing molecules from first principles to have minimal biological activity. Advances in computational chemistry and mechanistic toxicology provide the fundamental knowledge to advance the rational design of chemicals with minimal unintended biological activity. Although risk models are very useful in regulatory decision making, models that can characterize the intrinsic hazard of a chemical can be useful to practitioners of industrial ecology, toxicology, chemistry, and engineering.

Development of in silico methods for estimation of toxicity from chemical structures has advanced considerably in recent decades, with significant emphasis on quantitative structure–activity relationships (QSARs) (30, 31). However, predictive ability of QSARs is often hindered by model training issues, such as

---

Significance

Two of the rigorous disciplines that have emerged over the last 20 y to empower sustainability science are industrial ecology and green chemistry. Robust assessment tools of industrial ecology identify the greatest opportunities to mitigate human health and environmental impacts resulting from human activity. Green chemistry designs and develops chemicals, materials, and processes that, throughout the life cycle, minimize hazard and maximize efficiency. This process often entails synthesizing new molecules while maintaining function and minimizing adverse outcomes, particularly toxicity. There is an urgent need to develop accurate and economical screening tools that predict potential toxicity and inform the design of safer alternatives. A computational approach is presented for the rational design of molecules for reduced acute aquatic toxicity.

---

domain applicability, overtraining, and model bias (32). As a result, QSARs have not successfully replaced in vitro and in vivo testing for many endpoints (33, 34). Further, QSARs are not intended to directly inform chemical design, as they cannot be used to qualitatively assess whether a particular structural modification will result in a different toxicity profile. This information is critical for efficiently and effectively designing alternative chemicals and materials that mitigate toxicity risks across the life cycle.

The study presented here investigates and evaluates a possible approach to the development of a rapid screening tool based on design guidelines for property ranges. The approach differs from QSAR in that rather than predicting a toxicity value or a threshold, it elucidates the probability that a compound with particular properties will exhibit a certain toxicity profile. Thus, the approach does not yield deterministic values, but rather probabilistic information about whether a chemical is likely to exhibit toxicity above or below a certain threshold value. The output of this approach can help reduce in vitro and in vivo testing by using key properties to define a high-probability safe chemical space of low to no toxicity. Further, by using property-based criteria as opposed to QSAR descriptors, informed design decisions can more readily be pursued for modifications to reduce potential toxicity.

We demonstrated this property-based approach to screen and delineate a wide range of commercial chemicals from those that are listed on the Environmental Protection Agency's (EPA) Toxic Release Inventory (35). In the current study, we aim to designate guidelines based on relevant physiochemical properties to inform acute aquatic toxicity reduction using the fathead minnow as an indicator organism. This effort builds on earlier work aimed to identify boundary values of key properties for the screening of specific toxicity endpoints (36–38). Refining our previous approach, design guidelines presented here are based on properties related to bioavailability and reactivity. When implemented, 99% of compounds in the highest acute aquatic toxicity category, and 89% of those in the moderate category are excluded. Further efforts are made to understand the relationship of these properties to individual modes of action for acute aquatic toxicity and to validate this model for organisms at multiple trophic levels.

Recent LCAs of chemicals and chemical classes (39–43) clearly define opportunities to mitigate adverse outcomes through informed design of functionally equivalent molecules with reduced toxicity. The approach presented here, using acute aquatic toxicity as a case study, offers a rational tool to inform the design of these new chemicals. Of course, it is necessary to iterate between LCA and these design tools to ensure that environmental and human health burdens are not being traded off between impact categories or shifted elsewhere in the life cycle when these molecules are synthesized and implemented at scale.

Results

Categorization of Acute Aquatic Toxicity Data. Using an acute toxicity dataset for the fathead minnow (44), the 555 compounds (Table S1) were divided into four categories based on the mean lethal concentrations (LC_{50}) as defined by the US EPA (45). Molar units were used to minimize error in toxicity threshold values resulting from differences in molecular weight. Accordingly, the established EPA toxicity thresholds in mass units were also converted to molar units (*Computational Approaches for Identifying Chemicals with Minimal Acute Toxicity to the Fathead Minnow*). The resulting thresholds and categorizations are shown in Table 1. Note that category 4 was added to distinguish chemicals with ultralow risk (or effectively no risk) of acute aquatic toxicity.

Determination and Evaluation of Property Boundary Value Screens. The goal of this effort is to investigate and evaluate a possible

approach for the development of a rapid screening tool based on boundary limits for relevant properties. In the case of acute aquatic toxicity, the relevant properties relate to bioavailability parameter(s) and electronic reactivity parameter(s). Using discriminant analyses we identified two pairs of properties as being the most effective at probabilistically determining the acute aquatic toxicity category of a given chemical (Table 1). The first property pair is octanol-water partition coefficient ($\log P_{o/w}$) and the energy gap between the highest occupied and lowest unoccupied frontier orbitals (ΔE). The second property pair is octanol-water distribution coefficient at pH 7.4 ($\log D_{o/w}$) and ΔE. Note that the partition coefficient is a ratio of the concentrations of the non-ionized compound between two different phases, whereas the distribution coefficient is the ratio of the sum of the concentrations for both forms of the compound (ionized and nonionized). For unionizable compounds, $\log P_{o/w}$ is equal to $\log D_{o/w}$. To identify cutoff values a recursive algorithm in the R statistical environment was applied that maximized the ratio in Eq. **1**, which was adapted from ref. 46:

$$\text{Optimization Ratio} = \frac{\%\ \text{False positives}}{\%\ \text{False negatives}}. \qquad [1]$$

From Eq. **1**, false positives are defined as compounds in the low and no concern categories for acute aquatic toxicity that are inadvertently screened out because they do not meet the $\log D_{o/w} < X$ and $\Delta E > Y$ eV criteria, while false negatives are defined as compounds in the high and moderate concern categories that are retained because they meet the $\log D_{o/w} < X$ and $\Delta E > Y$ eV criteria; and X and Y are values of $\log D_{o/w}$ and ΔE, respectively, that were varied by 0.05 increments.

In Fig. 1*A*, compounds that are not of concern for acute aquatic toxicity (Table 1, category 4; colored green) are almost entirely confined to the quadrant of the plot defined by boundaries of $\log P_{o/w} < 1.7$ and $\Delta E > 6$ eV. Similar results were obtained when $\log D_{o/w}$ is substituted for $\log P_{o/w}$ (Fig. 1*B*). However, the optimized boundary value of $\log D_{o/w} < 1.7$ is able to confine 100% of the compounds in the no concern category (green), whereas only 93% are confined by the same optimized value for $\log P_{o/w}$; that is, a higher maximum optimization ratio (Eq. **1**) was realized. This result indicates that $\log D_{o/w}$ offers higher sensitivity than $\log P_{o/w}$ for identifying chemicals with no concern for acute aquatic toxicity. As such, $\log D_{o/w}$-ΔE, rather than $\log P_{o/w}$-ΔE, was used as the property pair for further analysis.

From Table 2, applying property boundary values for $\log D_{o/w}$ and ΔE sequentially as parameters for screening out chemicals leads to a decrease in the number of compounds in the high and moderate concern categories. At the same time, the mean LC_{50} value for the chemicals that are retained by these screens increases. (A third property boundary value identified in Table 2 is discussed below.) Using the property boundary value of 1.7 for $\log D_{o/w}$ retains 100% of the compounds in the no concern category; however, 12% and 27% of the compounds in the high and moderate concern categories, respectively, also remain. Overall,

Table 1. Thresholds and categorizations of acute toxicity levels of concern (45) using the molar mean lethal concentration (LC_{50}) for a dataset reporting fathead minnow 96-h acute toxicity (44)

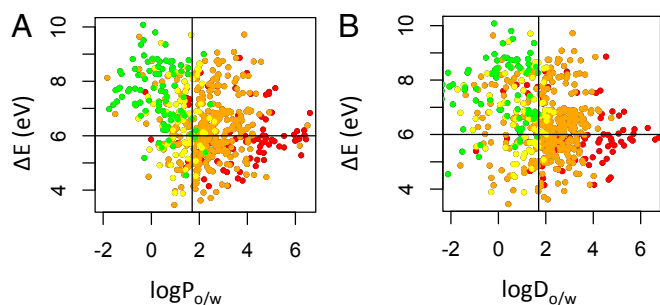| Category	Level of concern	Lower limit (LC_{50}, mmol/L)	Upper limit (LC_{50}, mmol/L)	Number of compounds
1	High	0	0.0067	81
2	Moderate	0.0067	1.49	305
3	Low	1.49	3.32	80
4	None	3.32	—	89
			Total	555

Fig. 1. Scatter plots of (*A*) the octanol-water partition coefficient (log$P_{o/w}$) vs. energy difference between the highest occupied and the lowest unoccupied frontier orbitals (Δ*E*) and (*B*) the octanol-water distribution coefficient (log$D_{o/w}$) vs. Δ*E*. The 555 compounds represented are colored by category of concern for acute aquatic toxicity as described in Table 1 (red, high concern; orange, medium concern; yellow, low concern; green, no concern) based on a 96-h toxicity assay of the fathead minnow (44).

this filter increases the mean LC$_{50}$ value to 2,247 mg/L or 1.29 mmol/L. Adding the second criterion, Δ*E* > 6 eV, reduces the percent of compounds in high, moderate, and low concern categories to 7%, 15%, and 48%, respectively, while still capturing 89% of the compounds in the no concern category. The second criterion also further increases the mean LC$_{50}$ value of the retained chemicals to 3,006 mg/L or 2.71 mmol/L.

To further improve the accuracy of the property-based filters, a third molecular property, molecular volume, was identified to screen out compounds retained in the high concern category after applying the log$D_{o/w}$ and Δ*E* criteria (Fig. 1*B*, red chemicals in upper left quadrant). The box plot in Fig. 2 shows the distribution of molecular volume (V) across the four categories of concern defined in Table 1. A recursive algorithm to maximize the ratio in Eq. **1** was used to determine a boundary value of V > 620 Å^3 using 5-Å^3 increments. Applying this additional property value screen leads to a negligible reduction in the chemicals in the no concern category being retained, while reducing the false-negative rate of the most toxic compounds from 7% to 1% (Table 1, bottom row). The upper limit of this volume-based design guideline can be defined as 920 Å^3 based on the current data; however, we recognize that this threshold is dependent on the effect of molecular volume on bioavailability and metabolism. Because a mechanistic threshold has not yet been determined to the best of our knowledge, it is feasible that the true threshold value could be higher than 920 Å^3.

In this safer chemical space defined by log$D_{o/w}$ <1.7, Δ*E* > 6 eV and V < 620 Å^3, the population of compounds in the moderate to low concern categories was decreased by 3–4%, and the mean LC$_{50}$ value of the subset increased to 3,405 mg/L or 3.65 mmol/L. With these filters, the only high concern compound in the dataset

that remained was allyl alcohol, which is postulated to undergo metabolic conversion to acrolein in the liver (47). Acrolein is more reactive and would be filtered from the safer chemical space using the Δ*E* > 6 eV rule, because the computed Δ*E* for acrolein is 5.86 eV.

Relationship with Toxic Modes of Action. Fig. 3 depicts the relationship between the property-based filters and the four largest classes by toxic modes of action (MOA) in the dataset: narcotics, polar narcotics, electrophiles or proelectrophiles, and compounds with undetermined MOAs. Additionally, Table 3 quantifies the impact of sequentially applying the property-based filters (log$D_{o/w}$ <1.7, Δ*E* > 6 eV, V < 620 Å^3) on categorizing chemicals into their experimentally determined MOAs. For narcotics, the property boundary of log$D_{o/w}$ < 1.7 eliminated nearly all of the high and moderate concern chemicals, with the remainder being excluded by Δ*E* > 6 eV. At the same time, 95% of the no concern chemicals were retained. A similar pattern was observed in the case of polar narcotics, where the log$D_{o/w}$ < 1.7 and Δ*E* > 6 eV filters substantially reduced the moderate and low concern compounds while retaining five of the six chemicals in the no concern category.

For electrophiles and proelectrophiles, log$D_{o/w}$ < 1.7 screened out 21 of the 23 high and 30 of the 61 moderate concern chemicals. Adding Δ*E* > 6 eV as an additional filter screened out 1 of the 2 high, 12 of the remaining 30 moderate, and 2 of the 8 low concern chemicals, retaining the single no concern chemical. In the case of the nondetermined MOA class, the volume criterion played a significant role by eliminating all 5 of the high and 11 of the 16 moderate concern chemicals while retaining 21 of 22 chemicals in the no concern category.

It is of interest to examine the compounds in the no concern category that are inadvertently screened out by the three property-based filters (false positives; Fig. S1), as well as the high and moderate concern compounds that are not filtered out (false negatives; Fig. S2). Only a few functional groups are represented in the false-negatives subset, and for many of these, such as esters, amides, primary amines, and allyl alcohol derivatives, metabolic transformations cannot be ruled out. Thus, future iterations to improve the accuracy of these property-based filters for identifying and designing chemicals with reduced acute aquatic toxicity should consider the role of metabolism, evaluating both the parent compounds and likely metabolites. Although frontier orbital energies give consistently good correlations with toxicity thresholds across all mechanisms, other electronic parameters also perform well within MOA subsets (Fig. S4 and Table S2). Some of these electronic parameters are further discussed in *Supporting Information*; they are complementary to frontier orbital energies, and can yield insight to the types of chemical interactions involved in the rate-determining steps of these MOAs.

Table 2. Percent of compounds that fall in each acute aquatic toxicity category of concern (Table 1) as different property filters are applied

Property-based filter	Acute aquatic toxicity concern category				Mean LC$_{50}$ of compounds retained	
	High (%)	Moderate (%)	Low (%)	None (%)	mg/L	mmol/L
None	15	55	15	15	999	0.155
log$D_{o/w}$ < 1.7	12	27	80	100	2,247	1.29
log$D_{o/w}$ < 1.7; Δ*E* > 6 eV	7	15	48	89	3,006	2.71
log$D_{o/w}$ < 1.7; Δ*E* > 6 eV; V < 620 Å^3	1	11	45	88	3,405	3.65

The mean LC$_{50}$ values (both mass and molar based) of the compounds retained by each property filter are also reported.
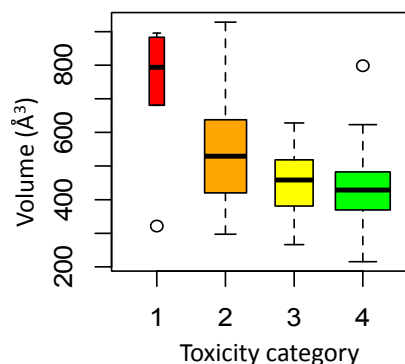
Fig. 2. Box plot showing distribution of molecular volume (V) by concern category for acute aquatic toxicity for compounds (Table 1) that have $\log D_{o/w} < 1.7$ and $\Delta E > 6$ eV.

External Validation of the Property-Based Design Guidelines. As with any statistical predictive approach, it is crucial to ascertain the degree of the tool's robustness to provide a user with confidence when making decisions about compounds not included in the training set. External validation was carried out using acute toxicity dataset of 345 compounds for a different aquatic species, a cladoceran (*Daphnia magna*) (48). This dataset is structurally diverse and, more importantly, only 15% of the compounds overlap with the fathead minnow dataset. Applying the property-based filters ($\log D_{o/w} < 1.7$, $\Delta E > 6$ eV, $V < 620$ Å^3) eliminated 95% of the compounds in the high and 86% of those in the moderate acute aquatic toxicity categories. At the same time, 55% and 67% were retained, respectively, in the low and none toxicity categories. The somewhat diminished ability to retain compounds in the no toxicity category (67% vs. 88% for fathead minnow) can be attributed to having only 12 compounds of this category in the validation dataset overall. Only five compounds in the high toxicity category were not filtered out, and of one them (allyl alcohol) was also an outlier in the training set. As discussed previously, this compound can be eliminated on the basis of metabolic conversion to a more reactive acrolein in the liver. Thus, the developed design guidelines have good general applicability, consistently showing that (*i*) nearly all high toxicity category compounds are excluded and (*ii*) increasingly more

compounds are eliminated proceeding from the none to high toxicity category with the application of the property-based filters. A scatter plot showing the ability of two of the criteria, $\log D_{o/w} < 1.7$ and $\Delta E > 6$ eV, to discern chemicals belonging to different acute aquatic toxicity categories is shown in Fig. 4.

Conclusions

One of the significant challenges in sustainability science is quantifying and minimizing the impact of production and consumption on human health and the environment. Synergistically, industrial ecology and green chemistry can make significant strides toward this goal through advancing rapid, accurate, economical, and robust assessment and design tools. Based on recent industrial ecology assessments of the chemical sector, there is significant potential to mitigate the human health and environmental impacts through the design of less toxic chemicals. Here, we demonstrated an approach using boundary values of relevant physiochemical properties to define a safer chemical space using acute aquatic toxicity as an example endpoint. By applying the proposed property-based filters on descriptors related to bioavailability and chemical reactivity as a screen, 99% of compounds in the high acute aquatic toxicity category and 89% of those in the moderate category are excluded, whereas 45% and 88% are retained, respectively, for the low and none toxicity categories. Further, the proposed filters were applied to a toxicity dataset for an aquatic organism at a different trophic level with similar results in terms of screening out compounds of high to moderate concern and retaining those with low to moderate concern for acute aquatic toxicity.

We also examined chemicals in the dataset by mode of action and found that octanol-water distribution coefficient and the energy gap between the highest occupied and lowest unoccupied molecular orbitals are effective at defining safer chemicals space for the largest MOA classes: narcotics, polar narcotics, and electrophiles/proelectrophiles. The molecular volume criterion improves the toxicity designation of compounds with nondetermined MOA. Our approach demonstrates the feasibility and potential to use property-based filters to identify chemicals of potential concern for a specific toxicity endpoint, enhance prediction of human health and environmental impacts, and enable the design safer chemicals through probabilistic information. Further, the models may potentially support industrial
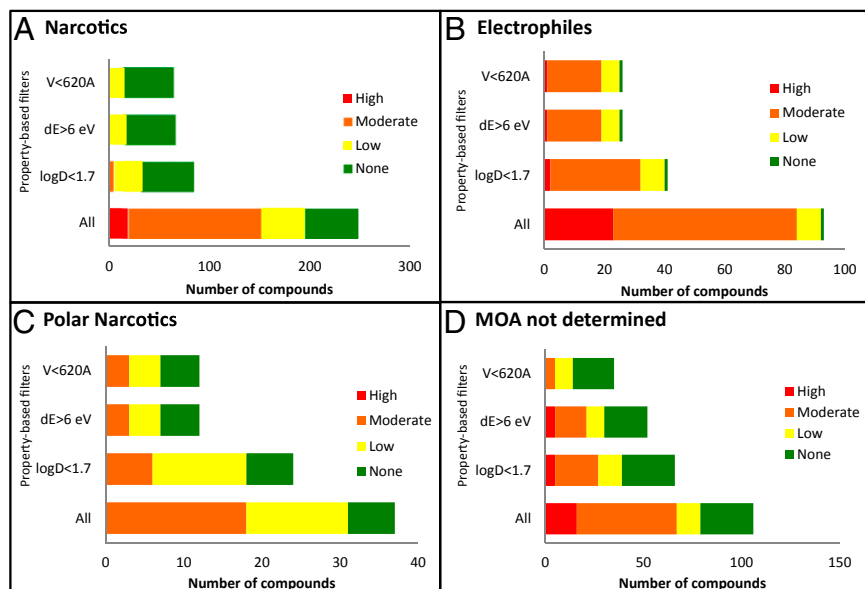

Fig. 3. Bar graphs showing distribution of the four toxicity concern categories (Table 1) as each property filter is applied for (*A*) narcotics, (*B*) electrophiles or proelectrophiles, (*C*) polar narcotics, and (*D*) compounds with undetermined toxic MOAs. log*D*, octanol-water distribution coefficient at pH 7.4; d*E* (eV), energy difference between the highest occupied and lowest unoccupied frontier orbitals; V, molecular volume (Å^3).

Table 3. Percent of compounds that fall in each experimentally determined MOA category when property filters are applied

Property-based filter	MOA									
	N	PN	ACE	UOP	CNS	RB	EP	NDP	ND	Total
None	274	37	16	12	9	2	93	6	106	555
$\log D_{o/w} < 1.7$	34	65	19	50	44	100	44	50	62	44
$\log D_{o/w} < 1.7$; $\Delta E > 6$ eV	27	32	6	0	11	100	28	17	49	30
$\log D_{o/w} < 1.7$; $\Delta E > 6$ eV; $V < 620$ Å^3	26	32	6	0	11	100	28	0	33	27

MOA categories are as follows: N, narcosis; PN, polar narcosis; ACE, acetylcholinesterase inhibition; UOP, uncoupler of oxidative phosphorylation; CNS, central nervous system seizure or stimulant; RB, respiratory blocker; EP, electrophiles; NDP, neurodepressant; ND, MOA could not be determined. log $D_{o/w}$, octanol-water distribution coefficient at pH 7.4; V, molecular volume (Å^3); ΔE (eV), energy difference between the highest occupied and lowest unoccupied frontier orbitals.

ecology assessments by generating more robust approximations of environmental impacts by a wide range of chemicals.

Methods

Toxicity Data and Validation. The acute toxicity dataset of 617 compounds for fathead minnow was obtained from 96-h EPA acute toxicity assays (44) reported in LC$_{50}$ values. The LC$_{50}$ values in milligrams per liter were transformed to millimoles per liter using the molecular weight of each chemical. Thirty-seven compounds that did not have reported LC$_{50}$ values and 25 inorganic compounds were omitted, reducing the total dataset to 555 compounds. The dataset is structurally diverse, consisting of 49 chemical classes. Table S1 lists the distribution of compounds within each chemical class. The acute aquatic toxicity for 345 compounds tested on Daphnia magna (LC$_{50}$ values at 48 h) was obtained from the Japanese Ministry of Environment (48).

Determining Octanol-Water Partition Coefficient. Pioneering work by Meyer and Overton first showed the relationship between activity of narcotics and their solubility in oils (49, 50). Others applied the concept of narcosis to aquatic toxicology, establishing the relationship between octanol-water partition coefficient (log$P_{o/w}$) and acute toxicity by narcosis (51–53). McFarland later showed that aquatic toxicity can be considered the result of penetration of toxicant into biophases and its interaction with one or more biochemical sites of action (54). Bioavailability of chemicals in aquatic organisms was shown to relate to absorption; for example, chemical partitioning across fish gills (55) represents an important route of exposure. Octanol-water partition coefficients were obtained from Russom et al. (44) and consisted of 218 measured values and 337 predicted values (CLOGP) (56). The univariate correlation of log(LC$_{50}$) values to log$P_{o/w}$ is high (−0.747), as anticipated based on the association of log$P_{o/w}$ with fish bioavailability and narcotic potency.
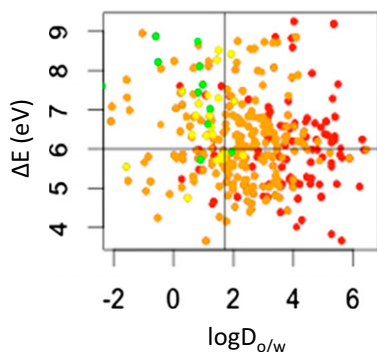

Fig. 4. Scatter plot of octanol-water distribution coefficient at pH 7.4 (log$D_{o/w}$) vs. the energy difference between the highest occupied and lowest unoccupied frontier orbitals. ΔE (eV) for 345 compounds, colored by category of concern for acute aquatic toxicity based on a 48-h toxicity assay of Daphnia magna (48). (red, high concern; orange, medium concern; yellow, low concern; green, no concern).

Determining Octanol-Water Distribution Coefficient at Biological pH. One shortcoming of log$P_{o/w}$ is that it does not take into account possible ionization of compounds, yet ionizability of organic chemicals strongly affects bioavailability to aquatic species (57, 58). Thus, it is more appropriate to consider log$D_{o/w}$, a distribution coefficient, which reflects the total concentrations of all species of a compound in the two phases at a given pH. In this study, log$D_{o/w}$ values were estimated at a biological pH of 7.4. It should be noted, however, that pH conditions often vary in aquatic assays and the natural aquatic environment and are rarely reported, which introduces an unavoidable error into our calculations. ChemAxon pK_a calculator plugin was used to predict pK_a and log$D_{o/w}$ values (Marvin v.6.0, 2013; ChemAxon). Each chemical class within the dataset has a wide distribution of log$D_{o/w}$ values as shown in Fig. S3.

Computational Approach. The importance of frontier orbitals on chemical reactivity is well known and is summarized in the frontier molecular orbital (FMO) theory pioneered by Fukui et al. (59). Energy separation between the highest occupied and lowest unoccupied molecular orbitals (HOMO–LUMO gap) has long served as a simple measure of kinetic stability. A molecule with a small HOMO–LUMO gap is considered chemically reactive for covalent bonding.

Our previously published procedure for computing HOMO–LUMO energies was revised to improve the accuracy of results by recognizing that different structural conformers can have different electronic properties. For example, even a small molecule like amino-2-propanol has 17 different conformers. Computed with B3LYP/6-31 + G(d) (49–52) in a polarizable continuum (IEF-PCM) (60, 61), these conformers span ∼3.7 kcal/mol in free energy and 10.5 kcal/mol (0.5 eV) in frontier orbital energies. Thus, Monte Carlo (MC) simulations in aqueous solution (TIP4P water model) (62) were first used to sample the conformational space of each compound in the dataset and locate the ground (i.e., lowest energy) states. Physiochemical properties considered as descriptors in this study, such as molecular volume, were computed as ensemble averages from MC simulations.

Reactivity parameters were derived from density functional theory (DFT). To this end, ground state conformers from MC simulations were optimized using the mPW1PW91 hybrid density functional (63) and the MIDIX+ basis set (64). The mPW1PW91/MIDIX+ approach was previously found to provide excellent performance-to-cost ratio for orbital energy calculations and can be readily used to evaluate larger molecules in reasonable time scales (65). In our study, HOMO energies calculated with mPW1PW91/MIDIX+ showed good qualitative agreement with experimental ionization energies (IEs) when applied to a diverse subset of 40 small organic molecules ($R^2 = 0.93$). Further, computed gap energies, ΔE, showed remarkable correlation ($R^2 = 0.99$) to results from B3LYP/6-31 + G(d) calculations, which can be considered a reasonable benchmark (66). In contrast, semiempirical methods (SMOs) performed significantly worse: AM1 (IE: $R^2 = 0.92$; ΔE: $R^2 = 0.78$), PDDG/PM3 (IE: $R^2 = 0.84$; ΔE: $R^2 = 0.68$), and PM6 (IE: $R^2 = 0.89$; ΔE: $R^2 = 0.76$). Poorer performance of SMOs can be attributed to parameterization and the use small basis sets when computing LUMO energies. Universal solvation model (SMD) (60) was used in all DFT calculations to estimate the influence of hydration.

MC simulations were carried out using BOSS 4.7 (67), and DFT calculations were performed using Gaussian g09 (68). The MIDIX+ basis set is not available as part of the Gaussian package and was downloaded from comp.chem.umn.edu.

Statistical Analysis. The R language and environment for statistical computing (69) (version 2.11.0) was used for exploratory data analysis, linear regressions, discriminant analyses, and identification of threshold values for properties that delineate safe chemical space. Descriptors that did not have normal distribution were transformed to a logarithmic scale.

1. Clark WC, Dickson NM (2003) Sustainability science: The emerging research program. *Proc Natl Acad Sci USA* 100(14):8059–8061.
2. Kates RW, et al. (2001) *Science* 292(5517):641–642.
3. Chen W-Q, Graedel TE (2012) Anthropogenic cycles of the elements: A critical review. *Environ Sci Technol* 46(16):8574–8586.
4. Du X, Graedel TE (2011) Global in-use stocks of the rare Earth elements: A first estimate. *Environ Sci Technol* 45(9):4096–4101.
5. Fischer-Kowalski M, et al. (2011) Methodology and indicators of economy-wide material flow accounting. *J Ind Ecol* 15(6):855–876.
6. Schaffartzik A, Eisenmenger N, Krausmann F, Weisz H (2014) Consumption-based material flow accounting. *J Industrial Ecol* 18(1):102–112.
7. Weisz H, Steinberger JK (2010) Reducing energy and material flows in cities. *Curr Opin Environ Sustainability* 2(3):185–192.
8. Wiedmann TO, et al. (2013) The material footprint of nations. *Proc Natl Acad Sci USA* 112:6271–6276.
9. Hellweg S, Fischer U, Scheringer M, Hungerbühler K (2004) Environmental assessment of chemicals: Methods and application to a case study of organic solvents. *Green Chem* 6(8):418–427.
10. Brunner PH (2012) Substance flow analysis. *J Ind Ecol* 16(3):293–295.
11. Eisler R (2010) *Handbook of Chemical Risk Assessment: Health Hazards to Humans, Plants, and Animals, Three Volume Set* (CRC Press, Boca Raton, FL).
12. Finnveden G, et al. (2009) Recent developments in life cycle assessment. *J Environ Manage* 91(1):1–21.
13. Pennington DW, et al. (2004) Life cycle assessment part 2: Current impact assessment practice. *Environ Int* 30(5):721–739.
14. Rebitzer G, et al. (2004) Life cycle assessment part 1: Framework, goal and scope definition, inventory analysis, and applications. *Environ Int* 30(5):701–720.
15. Biegler LT, Grossmann IE, Westerberg AW (1997) *Systematic Methods for Chemical Process Design* (Prentice Hall, Upper Saddle River, NJ).
16. Cano-Ruiz J, McRae G (1998) Environmentally conscious chemical process design. *Annu Rev Energy Environ* 23(1):499–536.
17. Chen H, Shonnard DR (2004) Systematic framework for environmentally conscious chemical process design: Early and detailed design stages. *Ind Eng Chem Res* 43(2):535–552.
18. Hoffmann VH, Hungerbühler K, McRae GJ (2001) Multiobjective screening and evaluation of chemical process technologies. *Ind Eng Chem Res* 40(21):4513–4524.
19. Sugiyama H, Fischer U, Hungerbühler K, Hirao M (2008) Decision framework for chemical process design including different stages of environmental, health, and safety assessment. *AIChE J* 54(4):1037–1053.
20. Anonymous (2005) *Sustainability in the Chemical Industry: Grand Challenges and Research Needs: A Workshop Report* (National Academies Press, Washington, DC).
21. Anonymous (2012) *Science for Environmental Protection: The Road Ahead* (National Academies Press, Washington, DC).
22. Anonymous (2013) *Public Health Linkages with Sustainability: Workshop Summary* (National Academies Press, Washington, DC).
23. Anonymous (2011) *Sustainability and the U.S. EPA* (National Academies Press, Washington, DC).
24. Judson RS, et al. (2010) In vitro screening of environmental chemicals for targeted testing prioritization: The ToxCast project. *Environ Health Perspect* 118(4):485–492.
25. Voutchkova AM, Osimitz TG, Anastas PT (2010) Toward a comprehensive molecular design framework for reduced hazard. *Chem Rev* 110(10):5845–5882.
26. Hauschild MZ, et al. (2013) Identifying best existing practice for characterization modeling in life cycle impact assessment. *Int J Life Cycle Assess* 18(3):683–697.
27. Rosenbaum RK, et al. (2008) USEtox—The UNEP-SETAC toxicity model: Recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment. *Int J Life Cycle Assess* 13(7):532–546.
28. Rosenbaum RK, et al. (2011) USEtox human exposure and toxicity factors for comparative assessment of toxic emissions in life cycle analysis: Sensitivity to key chemical properties. *Int J Life Cycle Assess* 16(8):710–727.
29. Rovida C, Hartung T (2009) Re-evaluation of animal numbers and costs for in vivo tests to accomplish REACH legislation requirements for chemicals: A report by the transatlantic think tank for toxicology (t(4)). *ALTEX* 26(3):187–208.
30. Cronin MT (2010) *Quantitative Structure–Activity Relationships (QSARs)–Applications and Methodology. Recent Advances in QSAR Studies* (Springer, New York), pp 3–11.
31. Cronin MT, et al. (2003) Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environ Health Perspect* 111(10):1391–1401.
32. Feher M, Ewing T (2009) Global or local QSAR: Is there a way out? *QSAR Comb Sci* 28(8):850–855.
33. Eriksson L, et al. (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect* 111(10):1361–1375.
34. Gramatica P (2007) Principles of QSAR models validation: Internal and external. *QSAR Comb Sci* 26(5):694–701.
35. Voutchkova AM, Ferris LA, Zimmerman JB, Anastas PT (2010) Toward molecular design for hazard reduction-fundamental relationships between chemical properties and toxicity. *Tetrahedron* 66(5):1031–1039.
36. Kostal J, Voutchkova-Kostal A, Weeks B, Zimmerman JB, Anastas PT (2012) A free energy approach to the prediction of olefin and epoxide mutagenicity and carcinogenicity. *Chem Res Toxicol* 25(12):2780–2787.
37. Voutchkova AM, et al. (2011) Towards rational molecular design: Derivation of property guidelines for reduced acute aquatic toxicity. *Green Chem* 13(9):2373–2379.
38. Voutchkova-Kostal AM, et al. (2012) Towards rational molecular design for reduced chronic aquatic toxicity. *Green Chem* 14(4):1001–1008.
39. Capello C, Wernet G, Sutter J, Hellweg S, Hungerbühler K (2009) A comprehensive environmental assessment of petrochemical solvent production. *Int J Life Cycle Assess* 14(5):467–479.
40. Wernet G, Conradt S, Isenring HP, Jiménez-González C, Hungerbühler K (2010) Life cycle assessment of fine chemical production: A case study of pharmaceutical synthesis. *Int J Life Cycle Assess* 15(3):294–303.
41. Wernet G, Hellweg S, Hungerbühler K (2012) A tiered approach to estimate inventory data and impacts of chemical products and mixtures. *Int J Life Cycle Assess* 17(6):720–728.
42. Wernet G, Mutel C, Hellweg S, Hungerbühler K (2011) The environmental importance of energy use in chemical production. *J Ind Ecol* 15(1):96–107.
43. Wernet G, Papadokonstantakis S, Hellweg S, Hungerbühler K (2009) Bridging data gaps in environmental assessments: Modeling impacts of fine and basic chemical production. *Green Chem* 11(11):1826–1831.
44. Russom CL, Bradbury SP, Broderius SJ, Hammermeister DE, Drummond RA (1997) Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). *Environ Toxicol Chem* 16(5):948–967.
45. US Environmental Protection Agency. Technical Overview of Ecological Risk Assessment. Available at http://www.epa.gov/oppefed1/ecorisk_ders/toera_analysis_eco.htm. Accessed February 20, 2014.
46. Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* 240(4857):1285–1293.
47. Serafini-Cessi F (1972) Conversion of allyl alcohol into acrolein by rat liver. *Biochem J* 128(5):1103–1107.
48. National Institute of Technology and Evaluation (2008) Ecotoxicity test results. Available at www.safe.nite.go.jp/english/sougou/doc/html/select_help_e.html#env_tox. Accessed December 1, 2012.
49. Overton E (1901) *Studien uber die Narkose zugleich ein Beitragzur allgemeinen Pharmakologie* (Gustav Fischer, Jena, Germany).
50. Meyer HH (1901) Zur theorie der alkoholnarkose. Der einfluss wechselnder temperature auf wirkungsstarke und Theilungscoefficient der narcotica. *Arch Exp Pathol Pharmacol* 46:338–346.
51. Könemann H (1981) Quantitative structure-activity relationships in fish toxicity studies. Part 1: Relationship for 50 industrial pollutants. *Toxicology* 19(3):209–221.
52. van Wezel AP, Opperhuizen A (1995) Narcosis due to environmental pollutants in aquatic organisms: Residue-based toxicity, mechanisms, and membrane burdens. *Crit Rev Toxicol* 25(3):255–279.
53. Veith GD, Call DJ, Brooke LT (1983) Structure toxicity relationships for the fathead minnow, *Pimephales promelas*: Narcotic industrial chemicals. *Can J Fish Aquat Sci* 40(6):743–748.
54. Oberg T (2004) A QSAR for baseline toxicity: Validation, domain of application, and prediction. *Chem Res Toxicol* 17(12):1630–1637.
55. McKim JM, Schmieder PK, Erickson RJ (1986) Toxicokinetic modeling of [C-14] pentachlorophenol in the rainbow trout (*Salmo gairdneri*). *Aquat Toxicol* 9(1):59–80.
56. Anonymous (1988) *CLOGP* (Pomona College, Claremont, CA), 3.4.
57. Zhao YH, Yuan X, Yang LH, Wang LS (1996) Quantitative structure-activity relationships of organic acids and bases. *Bull Environ Contam Toxicol* 57(2):242–249.
58. McCarty LS, Mackay D, Smith AD, Ozburn GW, Dixon DG (1993) Residue-based interpretation of toxicity and bioconcentration QSARs from aquatic bioassays: Polar narcotic organics. *Ecotoxicol Environ Saf* 25(3):253–270.
59. Fukui K, Yonezawa T, Shingu H (1952) A molecular orbital theory of reactivity in aromatic hydrocarbons. *J Chem Phys* 20:722–725.
60. Miertus S, Scrocco E, Tomasi J (1981) Electrostatic interaction of a solute with a continuum: A direct utilization of abinitio molecular potentials for the prevision of solvent effects. *Chem Phys* 55(1):117–129.
61. Tomasi J, Mennucci B, Cances E (1999) The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level. *J Mol Struct Theochem* 464(1-3):211–226.
62. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935.
63. Adamo C, Barone V (1998) Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The mPW and mPW1PW models. *J Chem Phys* 108(2):664–675.
64. Easton RE, Giesen DJ, Welch A, Cramer CJ, Truhlar DG (1996) The MIDI! basis set for quantum mechanical calculations of molecular geometries and partial charges. *Theor Chim Acta* 93(5):281–301.
65. Lynch BJ, Truhlar DG (2004) Small basis sets for calculations of barrier heights, energies of reaction, electron affinities, geometries, and dipole moments. *Theor Chem Acc* 111(2-6):335–344.
66. Zhan CG, Nichols JA, Dixon DA (2003) Ionization potential, electron affinity, electronegativity, hardness, and electron excitation energy: Molecular properties from density functional theory orbital energies. *J Phys Chem A* 107(20):4184–4195.
67. Jorgensen WL, Tirado-Rives J (2005) Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J Comput Chem* 26(16):1689–1700.
68. Frisch MJ, et al. (2009) *Gaussian 09 Revision A.1* (Gaussian, Wallingford, CT).
69. Team RDC (2009) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).