ACS CHEMICAL HEALTH & SAFETY

pubs.acs.org/acschas Article

Pesticide Indirect Photodegradation Database: A Data-Sharing Platform for Screening Existing and Discovering New Agrochemicals

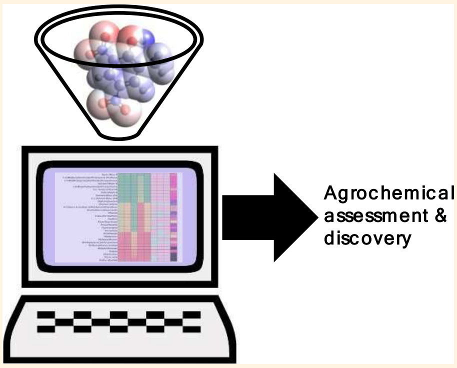
Zachary Stickelman, Natalie Clay, and Jakub Kostal*

Cite This: *ACS Chem. Health Saf.* 2023, 30, 420−427

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: The ability to control and tune physicochemical properties that underscore chemical behavior in living systems and the environment is at the "heart" of green chemistry. This is especially true for chemical classes designed a priori to be biologically active, such as pesticides, where the chance of unintended adverse outcomes is high. We recently proposed a design-vectoring framework, leveraging validated computational models of ecotoxicity and indirect photodegradation as a useful, quasisystems-based tool for screening existing and designing new agrochemicals. Here, we describe the development of a database that integrates our models, which link structural and substructural features to process metrics, and corresponding predicated data for all agrochemicals with photodegradable cores on the U.S. Environmental Protection Agency's registry (785 compounds and over 18,000 pairwise interactions with chromophoric dissolved organic matter, CDOM). The database is searchable by structural and nonstructural identifiers (e.g., chemical class, oxidizable core, physicochemical and electronic properties, etc.) to aid in chemical selection, hazard, and alternative assessment. Crucially, it can be easily updated and augmented to aid in interactive data-sharing across industry, government, and academia. The overarching goal of this project is to spur grander efforts in systems-based design of pesticides that would see this platform paired with target-based computational methods and incorporated into the discovery phase of new product development across industry sectors.



KEYWORDS: *pesticides, predictive modeling, degradation, alternative assessment, ecotoxicity*

■ INTRODUCTION

It is safe to say that there are no safe chemicals (pun intended), given that safety is freedom from risk and risk is a probability function, which can only be minimized but never abolished. Chemical risk is driven by hazard, which is encoded in the compound's molecular structure, with exposure acting as the confounding variable under conceivable use (and misuse) scenarios. Because the latter is difficult to characterize and much less quantify, unless we are concerned with exposure in a strictly controlled setting (e.g., a highly regulated chemical laboratory), the focus of risk assessment in new chemical development should be on hazard minimization, consistent with modern green chemistry practice.[1−4] This is a smart choice for chemical classes where toxicokinetic and toxicodynamic processes are mechanistically orthogonal to performance outcomes (e.g., azo dyes in textiles/printing, orthophthalates in food-contact applications, organohalogenated flame retardants, polyfluoroalkyl substances in oil/grease-resistant applications, etc.), and hazard can be gauged from relevant structure−activity relationships.[5,6] However, for chemicals with intended biological function (e.g., pesticides or pharmaceuticals), those pathways (i.e., target vs nontarget toxicity) may overlap to an extent that thwarts risk minimization and imposes inescapable trade-offs in new product development[7] (this is generally true for all chemical classes, where transformations that support function are mimicked in the living systems, generating toxicity—e.g., antioxidants and antiozonants used in car tires[8,9] or couplers used in peptide synthesis[10]). The pharmaceutical industry has grappled with this conundrum for decades, resulting in the development of rather comprehensive, systems-based frameworks that optimize the underlying trade-offs between function, adverse effects, and environmental persistence.[11] However, it can cost the pharmaceutical industry upward of 2.8 billion dollars to bring a new medicine to market,[12] reflecting the vast amount of testing and molecular engineering that takes place prior to commercialization. Most commodity chemicals have far less favorable (perceived) societal benefits, driving the need to develop new products at a much lower cost. Even active ingredients (AIs) in agricultural products, which are produced at volumes ca. 20-fold higher than medicinal products,[13−15]

Received: July 31, 2023
Revised: October 4, 2023
Accepted: October 11, 2023
Published: October 23, 2023

https://doi.org/10.1021/acs.chas.3c00076

cost only a fraction of pharmaceuticals in development and registration fees—low hundreds of millions of dollars on average.[16] This number can be seen both as too high to innovate frequently (compounded by rising unaffordability of agricultural products and their declining global cost−benefit ratio)[17] and too low to compete with the rigor of the pharmaceutical sector's development process to ensure high probability of safety and adequately characterized trade-offs.

One avenue for method translation from drug discovery to lower-cost chemical sectors is in the application of computational or in silico models, which are considered fast, inexpensive, and increasingly more accurate.[18,19] These methods are now a mainstay in every pharmaceutical company to help drive chemical design along multiple vectors of efficacy, safety, and depletion in the upstream of a new active pharmaceutical ingredient (API) discovery,[20] the latest prominent example being the COVID-19 treatment and vaccine development.[21,22] In the least, these tools help downsize the vastness of the chemical space into a manageable subset that has an increased probability of displaying certain (favorable) characteristics.[23−25] It is unclear to what extent these practices have already migrated to sectors such as pesticide design, given the secrecy of industrial practices to protect intellectual property. Anecdotal evidence suggests a wide gamut of capabilities, which, based on our conversations with industry insiders and regulators, range from "none at all" to "a decent replica of the computational drug discovery process." The one element that quite clearly appears to be missing, however, is a computational consideration of controlled and tunable depletion, which is concerning given increased risk to human health because of elevated exposure to persistent pesticides and/or their transformation products.[26]

Recently, we put forth a computational blueprint for controlling pesticides' net effects by exploiting physical−chemical mechanisms in the environment (i.e., soil, water, and air). Our approach to designing agrochemicals integrates existing, validated in silico strategies for environmental toxicity[27] with indirect photodegradation,[28] which is the main pathway for removing pesticides from the environment.[7,28,29] The goal was to construct a quasisystems-based framework for quantifying and reconciling trade-offs in the upstream of new product development (i.e., at the in silico stage of design vectoring toward a favorable part of the chemical space). For practical use, we defined cutoffs related to ecotoxicity and performance by mode of action (MOA) and pesticide class, which, when paired with color-coded photodegradation metrics, can be used to construct "design maps" to guide the (expert) end-user on molecular perturbations to achieve the desired outcome.[7]

Here, we outline the development of a more immediately useful data-sharing platform for the general (i.e., nonexpert) scientific community, which can be leveraged in hazard or alternative assessment using predicted metrics. Coined the pesticide indirect photodegradation (PIP) database, this live project is a collection of data on functionality, biologically relevant physicochemical properties, degradation kinetics and thermodynamics, core-substituent information (e.g., electron density distribution, orbital mixing, etc.), and environmental toxicity. Crucially, PIP integrates all our predictive models in a transparent manner, allowing users to interpret, edit, and expand the underlying data structure with relative ease. Our goal is to provide a "springboard" for interdisciplinary data-sharing[30] in safer pesticide development that can benefit the industry, regulators, and academics alike, with potential transferability to other chemical sectors.

■ METHODS

The PIP database (PIP.xlsm in Supporting Information, also available via download from https://kostal.columbian.gwu.edu/software/), was created by compiling data on 785 registered pesticides, which were obtained from the U.S. EPA's CompTox Chemical Dashboard and were categorized by pesticide class, function, and the type of oxidizable core (phenols, anilines, aryl ethers, sulfides, and thiols).[28,31] Predicted kinetic/thermodynamic information on indirect photodegradation with triplet-state chromophoric dissolved organic matter, ^{3}CDOM*, was generated based on published models[28] using experimental data in the public domain.[32−34] Our models calculate E_{cell} and log k (second-order rate constants) values from free energy barriers ($\Delta G^{\ddagger}$) and free energies ($\Delta G_{\text{et}}{}^0$) of the ^{3}CDOM*-pesticide electron transfer, respectively. Since both $\Delta G^{\ddagger}$ and $\Delta G_{\text{et}}{}^0$ can be estimated from the frontier molecular orbital theory (FMOT) using energies of the highest occupied molecular orbital (HOMO) of the pesticide and the singly occupied molecular orbital (SOMO) of the ^{3}CDOM*,[7,28] this information was incorporated into PIP along with corresponding linear models linking FMOT to $\Delta G^{\ddagger}/\Delta G_{\text{et}}{}^0$ and then to E_{cell}/log k. FMOT is a useful foundation for structure−property relationships and in PIP is further supported by Hirshfeld population (HPA) and natural bond orbital (NBO) analyses, which quantify substituent effects on pesticide oxidation potential ($\Delta G_{\text{et}}{}^0$) for select classes (phenols and anilines). The underlying raw data in PIP describe over 18,000 pairwise interactions with 23 ^{3}CDOM* components, which were grouped by the sensitizer class in the predictive models (e.g., ketones, aldehydes, quinones, coumarins, or PAHs−polycyclic aromatic hydrocarbons), as the specific composition of CDOM in nature is unknown.

Ecotoxicity metrics were integrated into PIP using the validated "Rule of Two" based on the density functional theory to describe the HOMO−LUMO, i.e., band gap (ΔE), which has long served as a useful predictor of covalent reactivity (vast majority of metabolic processes), and the octanol−water distribution coefficient (log $D_{\text{o/w}}$ at pH = 7.4), used to capture general bioavailability.[27,35,36] Past studies identified a safer chemical space (ΔE > 6 eV log $D_{\text{o/w}}$ < 1.7), which has a statistically higher probability of containing low-concern (i.e., benign) chemicals; for pesticides, we augmented this definition with Briggs rule (log $D_{\text{o/w}}$ < 3) to accommodate the function of AIs.[7] Finally, all pesticides were characterized by their pharmacological properties related to absorption, distribution, metabolism, and excretion (ADME), predicted from linear-response calculations based on Monte Carlo simulations in water using the BOSS program.[37] This approach to generate ADME properties has been successfully used in computational drug discovery,[38] which closely relates to pesticide AI design, both from the standpoint of safety and unintended hazard.[39] To support interpretation of ADME, we have documented relevant property ranges based on 95% confidence intervals for known drugs (viz., embedded MS Word file in PIP.xlsm). Color-coding across all developed parameters based on distribution percentiles was used to provide a facile semi-quantitative means of comparing different chemicals.

The development of PIP was carried out using Microsoft Excel VBA to facilitate accessibility and usability across fields and organizations. Functionality behind the data is built within the system as a self-sustaining database (i.e., PIP is perfectly usable in areas with limited internet). We anticipate that broad familiarity with MS Excel across scientific disciplines will support modifications to PIP (be it amendments to or expansion of existing data) based on the existing structure. This can in turn benefit the evolution of PIP's role in new product development beyond its current form, which focuses on alternative assessment in the upstream of agrochemical development. End-users can identify chemicals of interest and potential replacement analogs by leveraging the search and filter functions (coded using Microsoft Visual Basic on the backend). PIP supports selection of compound subsets based on single or multiple user-defined criteria to compare their metrics, which include exact/partial match for alphabetical strings (e.g., chemical names, classes, and SMILES) and value descriptors (e.g., equal/greater/smaller than or within specified range) for numerical strings.

■ RESULTS AND DISCUSSION

Database Structure and Intended Use. The intent for PIP is to offer guidance in the discovery phase of chemical development with regard to agrochemical photodegradation and ecotoxicity. Before discussing PIP's utility, it is important to stress that the underlying tools are probabilistic rather than deterministic though the latter methods can easily be incorporated in future iterations of PIP as outlined in our previous work.[6,18]

Relying on MS Excel was a strategic decision to build an easily accessible platform. The PIP.xlsm file contains five tabs. The first tab ("User Manual") explains the file structure and anticipated use and includes an embedded MS Word file that interprets predicted metrics and properties and outlines our methodology. The second tab ("All Pesticides") displays information about the primary function and a summary of experimental and modeled photodegradation and ecotoxicity data. This is the main tab for the end-user, where photodegradation data are reported for all three model tiers, i.e., substructural (NBO and HPA, for phenols and anilines only), structural (FMOT), and process ($\Delta G^{\ddagger}$, ΔG_{et}^{0} and corresponding log k, E_{cell}), as documented in our past work.[7,28] The third tab ("Data", Columns A-DU) is the complete data backend for the "All Pesticides" tab; thus, no data manipulation by the end user is allowed here. The second data table in this tab (Columns EA-EH), "Criteria", supports the user filtering process via Microsoft VBA. Finally, the third table (Columns EM-JG) contains data from the first table that fits the user's filter selections (i.e., data from the first table are copied to the third table following an advanced filtering process, which is subsequently displayed in the "All Pesticides" tab). The fourth tab ("Models"), describes all underlying models, partitioned by ^{3}CDOM* and pesticide class to offer additional refinement in alternative assessments. We recommend this tab for expert users familiar with the modeling protocol. The last tab ("Data1") is the source of all of the data displayed in the "Models" tab. Here, linear models developed by the Kostal group to predict $\Delta G^{\ddagger}$, ΔG_{et}^{0}, log k, and E_{cell} are embedded into every cell. The data structure of this tab is analogous to that of the "Data" tab.

The "All Pesticides" and "Models" tabs in PIP have two filtering features to optimize the functionality and user experience. The first filter consists of checkboxes on the left side of each of sheet. Given the overwhelming amount of raw data, this feature allows users to select the information they wish to display. A search bar at the top of the tab allows users to enter a specific query, such as a certain pesticide core, and a corresponding value range or a character string, to filter the entire data set to selections that match their criteria. The number of queries in this filter feature is unlimited, allowing users to progressively narrow data as necessary.

Interpreting outcomes are straightforward using Supporting Information in the "User Manual" tab; furthermore, color-coding is provided for all computed/predicted values based on distribution percentiles, which goes from red (unfavorable) to green (favorable) across the 785 pesticides in PIP. The user can thus lean on comparing specific values, color-coded outcomes, or "in" vs "out" based on predefined cutoffs. We should note here that photodegradation is reported on a continuous scale, while the ecotoxicity/function is based on value ranges. Despite the limited "resolution" of the latter, there is a precedent for statistically significant differences in toxicity outcomes when moving diagonally across the ΔE and log $D_{o/w}$-defined chemical space; our past work has shown that hazard generally increases going from low ΔE and high log $D_{o/w}$ toward high ΔE and low log $D_{o/w}$.[27]

Overview and Interpretation of Outcomes. In providing a basic overview of the 785 pesticide chemicals in PIP, only 52 compounds (or ca. 7% of the PIP database) meet the $\Delta E > 6$ eV and log $D_{o/w} < 1.7$ guidelines for safer chemicals (Figure 1), which is hardly surprising given that many of these compounds are biologically active ingredients. When factoring in the remaining criterion for safety, molecular volume (<620 Å),[27] only 12 chemicals meet these guidelines. One of these chemicals is the well-known herbicide dicamba, which is low in toxicity toward aquatic organisms, mammals, and honeybees, but is moderately toxic to birds.[40] Unfortunately, dicamba scores poorly in terms of indirect photodegradation (90th percentile in $\Delta G^{\ddagger}$ and ΔG_{et}^{0} values), owing to multiple electron-withdrawing rings on the aromatic core, which impede oxidation (i.e., the rate-determining electron transfer to ^{3}CDOM*). Because pesticides must be sufficiently lipophilic to exert selective toxicity, it is reasonable to amend this analysis, expanding the log $D_{o/w}$ cutoff (<3, as per Briggs Rule of 3), while maintaining the general reactivity criterion ($\Delta E > 6$ eV). This approach expands the list of compounds to 93; however, it now features compounds such as malathion (and

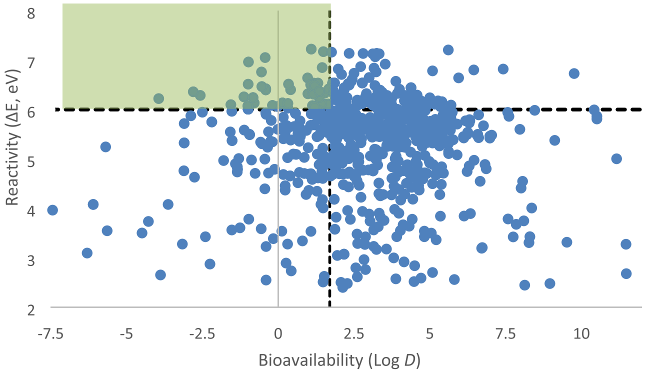


Figure 1. Density scatterplot of octanol−water distribution coefficient (log $D_{o/w}$) versus energy difference between the highest occupied and lowest unoccupied molecular orbitals (ΔE). Safer chemical space, defined by the current method (mPW1PW91/MIDIX+), is highlighted in the upper left-hand quadrant in green color (log $D_{o/w} < 1.7$ and $\Delta E > 6$ eV).

its reactive malaoxon metabolite) or oxamyl, which are both insecticides (acting as acetylcholinesterase inhibitors) with documented mammal toxicity and aquatic toxicity.[41] Thus, clear trade-offs emerge between function and toxicity when applying rather simple design guidelines.

Alternatively, one may consider decoupling the safety criteria (using either ΔE or log $D_{o/w}$), on the imperfect premise that, in the extremes, compounds that are not bioavailable can be reactive with no adverse effects and nonreactive compounds can be readily bioavailable due to lack of target-based toxicity. We note these scenarios since they are frequently invoked in green chemistry and safer chemical design,[42] and they can be problematic. While a lack of bioavailability may be a useful safety metric for many commercial chemicals, pesticide AIs rely on bioavailability to support selective toxicity; furthermore, highly reactive chemicals can cause adverse effects via dermal contact (e.g., irritation and sensitization). Conversely, nonreactive compounds may be metabolically activated into potent toxicants (viz. oxidation of malathion to malaoxon) or may be toxic via narcosis (i.e., disruption of ion channels by embedding into cell membranes and causing their swelling). The latter case appears more concerning than the former as documented in our previous analyses.[27,35,36] Considering the "off-diagonal" regions in Figure 1 (i.e., the lower-left and upper-right quadrants), there are 194 compounds that follow log $D_{o/w} < 1.7$ and 168 compounds that fit $\Delta E > 6$ eV. The distribution of oxidizable cores within these categories is shown in Figure 2. Most of the compounds that are considered safe in at least one parameter are aryl ethers or sulfides, which suggests that de novo pesticide design is needed primarily for the other two classes, i.e., phenols and anilines.

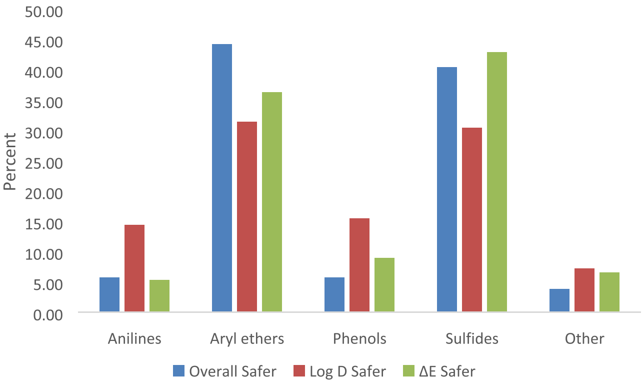


Figure 2. Relative representation of pesticide type (defined by oxidizable core) within each category of greater safety (overall safer = log $D_{o/w} < 1.7$ and $\Delta E > 6$ eV; log D safer = log $D_{o/w} < 1.7$; ΔE safer = $\Delta E > 6$ eV).

The limitation of the current hazard approach, which imposes clear trade-offs between function and toxicity for AIs can be alleviated in three ways: one, using performance-defined cutoffs (described in Lewer et al.),[7] which allow for some flexibility in structural manipulations that affect log $D_{o/w}$ and ΔE to both promote safety and retain the acceptable function for a given pesticide class and/or MOA; two, by promoting depletion as implemented in PIP, which assumes oxidation byproducts are of lesser concern due to increased water solubility; and three, by incorporating target-based models of toxicity, which offer greater level of refinement and selectivity.[6] Controlling depletion to mitigate hazard concerns is an interesting proposition. In PIP, this prediction rests on the pesticide's oxidation potential (i.e., electron-transfer to a ^{3}CDOM* mixture), which is reasonable given that reduction potentials of known ^{3}CDOM* sensitizers encompass other photochemically produced reactive intermediates (PPRIs), such as a singlet oxygen or a hydroxyl radical.[43] Furthermore, we assume that the electron transfer is the rate-determining step, followed by a downhill proton transfer to generate the pesticide radical.[33] While degradation pathways can yield (reactive) intermediate byproducts of concern, the end-products are usually of lower toxicity. For example, phenols and anilines are oxidized to carboxylic acids and eventually carbon dioxide and water[44] though semiquinone and quinone intermediates generated in these pathways can cause adverse effects (e.g., oxidative stress). Certain substitutions on oxidizable cores can impede detoxification transformations (most notably halogens), and the current models in PIP consider these effects in the initial electron transfer (e.g., halogens destabilizing the electron hole in the pesticide cation radical, resulting in less favorable thermodynamics).[7] Finally, the added benefit of abiotic (vs biotic) degradation models in PIP is that reactive intermediates, though potentially toxic, are less bioavailable than their parent molecules, which, paired with their kinetic instability, limits their risk to living systems. Accepting all the above, we must note the "crosstalk" between underlying models for ecotoxicity and photodegradation, which both rest on FMOT in PIP. To this end, we recommend the users lean on the energies of pesticides' LUMO for guidance on relative safety (in place of ΔE), given that most toxicants are electrophiles, while the HOMO energy plays a pivotal role in the pesticide photodegradation (i.e., oxidation). This approach allows one to "decouple" the two analyses in cases where there is too much overlap.[7]

Alternative Assessment Case Studies. Our general recommendations for alternative assessment follow the guidance outlined above, that is, to (i) use prescribed cutoffs ("in" vs "out") for ecotoxicity/performance where possible and (ii) use distribution percentiles (i.e., PIP color-coding) for photodegradation metrics as well as for ecotoxicity/performance when the use of boundaries is ineffective (i.e., selected compounds are all within the same chemical subspace). We also recommend consulting the pharmacological properties and their ranges to assess the biological activity with regard to ADME.

To demonstrate PIP's alternative assessment capabilities, we selected 3 compounds that perform poorly in one of the categories outlined in Figure 2: fails log D safety (2-ethylhexyl diphenyl phosphate), fails ΔE safety (formetanate), and fails overall safety (azafenidin). Azafenidin (herbicide aryl ether) is a herbicide and a protoporphyrinogen oxidase inhibitor, used to control weeds in fruit crops; 2-ethylhexyl diphenyl phosphate is a cholinesterase-inhibiting pesticide ingredient that is commonly used as a flame retardant and a plasticizer; and formetanate is a carbamate pesticide, its primary MOA being cholinesterase inhibition (Figure 3). The narrative below describes three alternative assessment case studies, each example selecting one of the chemicals in Figure 3 and then assessing the top alternatives with the same pesticide class and oxidizable core.

With computed log $D_{o/w}$ of 3.5 and a ΔE of 5.59 (56th percentile for LUMO energy), azafenidin is of concern, which is consistent with available safety data sheets (SDSs) that consider azafenidin to be moderately toxic to birds and highly

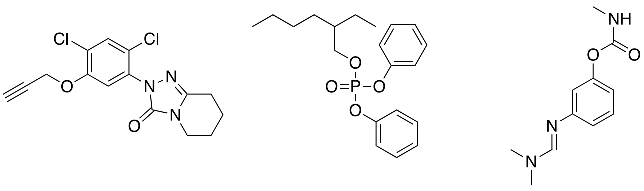


Figure 3. Compounds used in the 3 case studies. Left to right: azafenidin, 2-ethylhexyl diphenyl phosphate, and formetanate.

toxic to aquatic organisms.[41] In the consideration of alternatives, it is crucial to maintain the herbicide functional class and the aryl ether core type. Provided these parameters, PIP outputs the top 5 aryl ether herbicides, which are (*R*)-mecoprop, dicamba-dimethylammonium, dichlorprop, dicamba, and sodium dicamba. All of these compounds have ΔE above the 80th percentile and log $D_{o/w}$ below the 10th percentile. The top compound, sodium dicamba, has an ΔE of 6.24 eV (88th percentile in ΔE and 60th percentile in LUMO energy) and log $D_{o/w}$ of −0.83 (Figure 4). Both metrics are within the safer region in terms of ecotoxicity, and, if one accepts alternate vertebrates as a useful new approach methodology (NAM) for higher species,[4] also safe in terms of acute mammalian toxicity. The trade-off is poor photodegradation—90th percentile in computed $\Delta G^{\ddagger}$ and ΔG_{et}^{0} values, where higher values indicate a lower kinetic and thermodynamic propensity to undergo oxidation, respectively. To that end, (*R*)-mecoprop may be the better overall alternative (38th percentile in photodegradation $\Delta G^{\ddagger}$ and ΔG_{et}^{0} values, ninth percentile in log $D_{o/w}$, 73rd percentile in ΔE, and 77th percentile in LUMO energy). Consistent with our safety predictions, mecoprop is of low concern to aquatic species.[41]

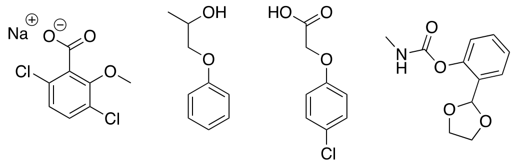


Figure 4. Alternative compounds identified for each case study in Figure 3 are based on the highest-predicted safety criteria. The compound on the left, sodium dicamba, is an alternative to azafenidin; the two compounds in the middle, 4-chlorophenoxyacetic acid and 1-phenoxy-2-propanol, are alternatives to 2-ethylhexyl diphenyl phosphate; and the compound on the right, dioxacarb, is an alternative to formetanate.

The compound 2-ethylhexyl diphenyl phosphate (Figure 3) is an aryl ether with computed log $D_{o/w}$ and ΔE values of 6.45 and 6.80 eV, respectively. While its reactivity is within the safe cutoff value, it fails in predicted bioavailability (both based on our Rule of Two, log $D_{o/w} < 1.7$ and the Briggs Rule of Three, log $D_{o/w} < 3$). The top 5 outputs for the same pesticide class and oxidizable core were 1-phenoxy-2-propanol, 2-(2-(2-(2-phenoxyethoxy)ethoxy)ethoxy)ethanol, 3-hydroxycarbofuran, 4-chlorophenoxyacetic acid, and entsufon sodium. All of these compounds had ΔE above the 80th percentile and log $D_{o/w}$ below the 30th percentile. Depending on user preferences, two top choices were identified: 4-chlorophenoxyacetic acid, which has a particularly low log $D_{o/w}$ (−1.56) and an acceptable ΔE of 6.10 eV (79th percentile for LUMO energy), and 1-phenoxy-2-propanol, which is the least reactive alternative in the series (ΔE = 6.34 eV, 94th percentile in LUMO energy), with a predicted log $D_{o/w}$ (1.54) just below the prescribed cutoff (Figure 4). Both compounds perform reasonably well for photodegradation (51st and 42nd percentiles in computed $\Delta G^{\ddagger}$ and ΔG_{et}^{0} values, respectively), which compares favorably against 2-ethylhexyl diphenyl phosphate (91st percentile). From available SDSs, 4-chlorophenoxyacetic acid is moderately toxic to birds, honeybees, and some aquatic organisms; however, it has a relatively low toxicity to fish (our ecotoxicity model is dominated by the highly curated U.S. EPA fathead minnow data).[41,45] In contrast, 1-phenoxy-2-propanol is of low acute aquatic toxicity (LC_{50} = 280 mg/L in Fathead minnow, 370 mg/L in *Daphnia magna*, and >100 mg/L in algae);[41] the outcome difference between the two compounds being consistent with predicted relative ΔE and LUMO energy values.

The final example is a reverse scenario where an aryl ether insecticide, formetanate, meets the safe-space requirement for log $D_{o/w}$ (−0.25) but fails to meet the cutoff for ΔE (5.36 eV). From available SDSs, formetanate is moderately to slightly toxic to freshwater fish but highly (acutely) toxic to freshwater invertebrates.[41] Four alternatives for this core category and functional class were identified in PIP: dioxacarb, fenamiphos sulfone, fenamiphos sulfoxide, and bendiocarb, which all scored in the 20th percentile for log $D_{o/w}$ and above the 80th percentile for ΔE. Since all four alternatives met the safety guidelines, ranking by photodegradation is useful, where bendiocarb was found to be the easiest to oxidize (22nd percentile), followed by fenamiphos sulfoxide (67th percentile), dioxacarb (91st percentile), and fenamiphos sulfone (94th percentile). Here we should reiterate that our tools are probabilistic (vs deterministic) as neither of these compounds can be confidently identified to be of low concern to aquatic organisms based on available data. For example, bendiocarb is considered at least moderately toxic to several species of fish and aquatic invertebrates, such as shrimp, crabs, and insects;[41] dioxacarb was found moderately toxic to *Oncorhynchus mykiss* (rainbow trout),[41] which is an aquatic species not represented in our model, and is more sensitive than fathead minnow for certain chemical classes.[46]

Strengths, Limitations, and Future Improvements. When reflecting on the value of PIP, it is hard not to see the horizontal and democratic structure of our shared scientific enterprise as a detriment that leads to staggering inefficiencies.[30] In many ways, a "dictatorial" and hierarchical system in scientific data sharing would better systematize and organize all available information around a single standard, joining smaller-scope efforts (such as PIP or PPDB, the Pesticide Properties Database)[41] with larger endeavors (e.g., the U.S. EPA's CompTox Chemicals Dashboard or The Integrated Chemical Environment developed by NICEATM)[47] into a single platform. That is not to say there is no value in PIP as a standalone tool, but its utility is mostly as part of a weight-of-evidence scheme along with other sources of information. While combining multiple pieces of evidence to make a decision is standard practice in sciences, updating and expanding data, which is necessary on a continuous basis, is far more complicated with numerous data streams vs a single platform developed around a unified standard of data-management practices. To address this issue with PIP, we built this platform as an open-access, standalone tool, which can be readily revised and augmented by anyone familiar with the data types and MS Excel, and then disseminated to the rest of the scientific community. Undoubtedly, such "crowdsourcing" can be chaotic, but our hope is that "natural selection" in the hands of the scientific community ensures "survival of the fittest", i.e., helps PIP evolve to better serve the needs of the pesticides industry and its regulators.

In commenting on the limitations of PIP's structure and its underlying models, one must recognize both deficiencies of the in silico component and the underlying (experimental) data. We laid bare the assumptions and approximations used to assess indirect photodegradation and ecotoxicity in PIP, here and elsewhere,[7,27,28] and the end-user should exercise judgment in their use scenarios. By the same token, we should be cognizant of the experimental data variability, uncertainty, and reproducibility. Due to persisting negative bias toward in silico models, experimental outcomes (i.e., acute LC_{50} or chronic NOEC/LOEC values) are favored; however, that may be misleading if these data points are of low reliability (and have thus been excluded from model training). To that end, it is not uncommon to see several orders of magnitude variations in reported acute and chronic aquatic toxicity thresholds.[48] Data variability can occur because the chemicals are highly toxic or volatile or were tested close to their limit of solubility, in addition to reasons related to inter- and intraspecies differences, evolving test standards, experimental conditions, etc. The totality of these concerns generates significant uncertainty in any single data point.[49,50] Since high data quality is critical for robust model development,[51] one will inevitably sacrifice the breadth of the chemical space used for training, particularly across nonstandard test species, potentially putting model predictions at odds with those experimental outcomes.

Finally, any model ages and new data need to be interrogated on a continuous basis to test and correct existing trends and relationships. However, computational models are not always updated in such a way, particularly if they were funded by academic grants and are not used commercially. The current rule-based ecotoxicity model in PIP was developed in 2015 using the highly curated fathead minnow data (555 chemicals)[52] and was externally validated on a different aquatic species, a cladoceran (*Daphnia magna*).[27] Importantly, analogous definitions of safety in terms of log $D_{o/w}$ and ΔE cutoffs were noted for Japanese medaka and green algae, and held true in both acute and chronic toxicity testing, suggesting broad applicability of these mechanistic guidelines.[35,36] In 2016, the rule-based model became a foundation for the commercial computer-aided discovery and redesign (CADRE) tool, which predicts the U.S. EPA's categories of concern as well acute LC_{50}'s and chronic NOECs for fish, crustaceans, and algae using linear models based on physicochemical and electronic properties. CADRE is a reliable performer and is regularly updated to support testing needs of various industries.[53] In partnership with Oekotoxzentrum-center ecotox (https://www.ecotoxcentre.ch/),[54] the program predicted potency categories of 30 pesticides external to its training set with over 82% accuracy, suggesting the model can be used to resolve confounding cases encountered in PIP's alternative assessments. Alternatively, PIP's ecotoxicity predictions can be supplemented with open-access tools such as ECOSAR (included in the EPA EPISuite program and the OECD Toolbox) or TEST, which are provided by the U.S. EPA.[55,56] An overview of the robustness of these models is available elsewhere.[53]

Looking onward, the current structure of PIP can be amended to provide support beyond alternative assessment to help guide design of novel pesticides. This line of thinking is supported by our recent report, which illustrates how the substructural tier of our photodegradation model can be leveraged to control for depletion via rational modification of the oxidizable core substituents.[7] Since these HPA and NBO analyses are quantitative, they can be integrated into PIP on the backend to provide group (i.e., substituent) contributions to photodegradation $\Delta G^{\ddagger}$ and $\Delta G_{et}{}^{0}$ values per position on a given oxidizable core. In turn, the user can leverage this information to iteratively identify the most favorable perturbations of the basic molecular scaffold.

■ CONCLUSIONS

Understanding the factors that control safety, depletion, and performance of agrochemicals is vital in assessing alternatives that optimize the underlying trade-offs. Such an approach aligns with the recent emphasis on systems-thinking in green chemistry when designing safer and efficacious chemicals. Here, we describe the development of an open-access PIP database, which implements our previously validated models for indirect photodegradation and ecotoxicity. PIP is based on compiled experimental and computational data for 785 pesticides obtained from the U.S. EPA's registry (all pesticides with oxidizable cores) and is structured to provide support for alternative assessments of agrochemicals in the upstream of new product development.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chas.3c00076.

PIP database Microsoft Excel VBA file (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Jakub Kostal − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052-0066, United States;* orcid.org/0000-0001-9727-0477; Email: jkostal@gwu.edu

Authors

Zachary Stickelman − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052-0066, United States;* orcid.org/0000-0002-3131-2993

Natalie Clay − *Department of Chemistry, The George Washington University, Washington, District of Columbia 20052-0066, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.chas.3c00076

Author Contributions

The manuscript was written through the contributions of all authors. All authors have given approval for the final version of the manuscript.

Funding

This work was supported by NSF grant NSF1943127.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We would like to express our appreciation to G. Devineni and P. Griffin for their assistance with modeling.

REFERENCES

(1) Anastas, N. D. Connecting toxicology and chemistry to ensure safer chemical design. *Green Chem.* **2016**, *18* (16), 4325−4331.

(2) Constable, D. J. C. Green and sustainable chemistry - The case for a systems-based, interdisciplinary approach. *iScience* **2021**, *24* (12), No. 103489.

(3) Zimmerman, J. B.; Anastas, P. T. Toward substitution with no regrets. *Science* **2015**, *347* (6227), 1198−1199.

(4) Brooks, B. W.; Sabo-Attwood, T.; Choi, K.; Kim, S.; Kostal, J.; LaLone, C. A.; Langan, L. M.; Margiotta-Casaluci, L.; You, J.; Zhang, X. W. Toxicology Advances for 21st Century Chemical Pollution. *One Earth* **2020**, *2* (4), 312−316.

(5) DeVito, S. C. General principles for the design of safer chemicals: Toxicological considerations for chemists. *Des. Safer Chem.* **1996**, *640*, 16−59.

(6) Clymer, T.; Vargas, V.; Corcoran, E.; Kleinberg, R.; Kostal, J. Redesigning hazardous chemicals by learning from structure-based drug discovery. *Green Chem.* **2019**, *21* (8), 1935−1946.

(7) Lewer, J. M.; Stickelman, Z. R.; Huang, J. H.; Peloquin, J. F.; Kostal, J. Structure-to-process design framework for developing safer pesticides. *Sci. Adv.* **2022**, *8* (13), No. abn2058.

(8) Seiwert, B.; Nihemaiti, M.; Troussier, M.; Weyrauch, S.; Reemtsma, T. Abiotic oxidative transformation of 6-PPD and 6-PPD quinone from tires and occurrence of their products in snow from urban roads and in municipal wastewater. *Water Res.* **2022**, *212*, No. 118122.

(9) Tian, Z.; Zhao, H.; Peter, K. T.; Gonzalez, M.; Wetzel, J.; Wu, C.; Hu, X.; Prat, J.; Mudrock, E.; Hettinger, R. A ubiquitous tire rubber-derived chemical induces acute mortality in coho salmon (vol 9, pg 140, 2022). *Science* **2022**, *375* (6582), 185−189.

(10) Graham, J. C.; Trejo-Martin, A.; Chilton, M. L.; Kostal, J.; Bercu, J.; Beutner, G. L.; Bruen, U. S.; Dolan, D. G.; Gomez, S.; Hillegass, J. An Evaluation of the Occupational Health Hazards of Peptide Couplers. *Chem. Res. Toxicol.* **2022**, *35* (6), 1011−1022.

(11) Smith, R. E. *Systems thinking in medicine and new drug discovery*; Cambridge Scholars Publishing, 2018.

(12) Wouters, O. J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009−2018. *JAMA* **2020**, *323* (9), 844−853.

(13) Toma, A.; Crisan, O. Regulations on Green Pharmacy in European Countries - a Comparative Study. *Farmacia* **2021**, *69* (1), 182−188.

(14) Food and Agriculture Organization of the United Nations. *Pesticides use, pesticides trade and pesticides indicators*. https://www.fao.org/3/cc0918en/cc0918en.pdf (accessed October 2023).

(15) Toma, A.; Crişan, O. Green pharmacy - a narrative review. *Clujul Med.* **2018**, *91* (4), 391−398.

(16) Unglesbee, E. The Price of New Chemistry. In *Progressive Farmer*; DTN, 2016.

(17) Zhang, W. J. Global pesticide use: Profile, trend, cost/benefit and more. *Proc. Int. Acad. Ecol. Environ. Sci.* **2018**, *8* (1), 1−27.

(18) Kostal, J.; Voutchkova-Kostal, A. Going All In: A Strategic Investment in In Silico Toxicology. *Chem. Res. Toxicol.* **2020**, *33* (4), 880−888.

(19) Kostal, J.. Computational Chemistry in Predictive Toxicology: Status Quo et Quo Vadis? In *Advances in Molecular Toxicology*; Fishbein, J.. Ed.; Elsevier, 2016; vol *10*, pp 139−186.

(20) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303* (5665), 1813−1818.

(21) Chowdhury, U. F.; Shohan, M. U. S.; Hoque, K. I.; Beg, M. A.; Siam, M. K. S.; Moni, M. A. A computational approach to design potential siRNA molecules as a prospective tool for silencing nucleocapsid phosphoprotein and surface glycoprotein gene of SARS-CoV-2. *Genomics* **2021**, *113* (1), 331−343.

(22) Wang, J. M. Fast Identification of Possible Drug Treatment of Coronavirus Disease-19 (COVID-19) through Computational Drug Repurposing Study. *J. Chem. Inf. Model* **2020**, *60* (6), 3277−3286.

(23) Ferreira, L. G.; Dos Santos, R. N.; Oliva, G.; Andricopulo, A. D. Molecular docking and structure-based drug design strategies. *Molecules* **2015**, *20* (7), 13384−13421.

(24) Jorgensen, W. L. Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009**, *42* (6), 724−733.

(25) Vedani, A.; Smiesko, M. In Silico Toxicology in Drug Discovery - Concepts Based on Three-dimensional Models. *Atla-Altern. Lab. Anim.* **2009**, *37* (5), 477−496.

(26) Damalas, C. A.; Eleftherohorinos, I. G. Pesticide Exposure, Safety Issues, and Risk Assessment Indicators. *Int. J. Environ. Res. Public Health* **2011**, *8* (5), 1402−1419.

(27) Kostal, J.; Voutchkova-Kostal, A.; Anastas, P. T.; Zimmerman, J. B. Identifying and designing chemicals with minimal acute aquatic toxicity. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (20), 6289−6294.

(28) Lewer, J.; Huang, J.; Peloquin, J.; Kostal, J. Structure-Energetics-Property Relationships Support Computational Design of photodegradable Pesticides. *Environ. Sci. Technol.* **2021**, *55* (17), 11713−11722.

(29) 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.

(30) Kostal, J.; Brooks, B. W.; Smith, C. A.; Devineni, G. O data, where art thou? Revolutionizing data sharing to advance our sustainability goals through smart chemical innovation. *iScience* **2022**, *25* (18), No. 105256.

(31) McNeill, K.; Canonica, S. Triplet state dissolved organic matter in aquatic photochemistry: reaction mechanisms, substrate scope, and photophysical properties. *Environ. Sci.-Process. Impacts* **2016**, *18* (11), 1381−1399.

(32) Canonica, S.; Hellrung, B.; Wirz, J. Oxidation of phenols by triplet aromatic ketones in aqueous solution. *J. Phys. Chem. A* **2000**, *104* (6), 1226−1232.

(33) Canonica, S.; Jans, U.; Stemmler, K.; Hoigne, J. Transformation Kinetics of Phenols in Water - Photosensitization by Dissolved Natural Organic Material and Aromatic Ketones. *Environ. Sci. Technol.* **1995**, *29* (7), 1822−1831.

(34) Pavitt, A. S.; Bylaska, E. J.; Tratnyek, P. G. Oxidation potentials of phenols and anilines: correlation analysis of electrochemical and theoretical values. *Environ. Sci. Proc. Impacts* **2017**, *19* (3), 339−349.

(35) Voutchkova-Kostal, A. M.; Kostal, J.; Connors, K. A.; Brooks, B. W.; Anastas, P. T.; Zimmerman, J. B. Towards rational molecular design for reduced chronic aquatic toxicity. *Green Chem.* **2012**, *14* (4), 1001−1008.

(36) Voutchkova, A. M.; Kostal, J.; Steinfeld, J. B.; Emerson, J. W.; Brooks, B. W.; Anastas, P.; Zimmerman, B. Towards rational molecular design: derivation of property guidelines for reduced acute aquatic toxicity. *Green Chem.* **2011**, *13* (9), 2373−2379.

(37) Jorgensen, W. L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.* **2005**, *26* (16), 1689−1700.

(38) Duffy, E. M.; Jorgensen, W. L. Prediction of properties from simulations: Free energies of solvation in hexadecane, octanol, and water. *J. Am. Chem. Soc.* **2000**, *122* (12), 2878−2888.

(39) Zhao, W.; Huang, Y.-Q.; Hao, G.-F. Pesticide informatics expands the opportunity for structure-based molecular design and optimization. *Adv. Agrochem.* **2022**, *1* (2), 139−147.

(40) Minnesota Department of Agriculture. *Dicamba - GENERAL INFORMATION*. https://www.mda.state.mn.us/Dicamba-general-information#:~:text=Dicamba%20toxicity%20is%20low%20for,be%20damaged%20by%20this%20herbicide (accessed October 2023).

(41) Lewis, K. A.; Tzilivakis, J.; Warner, D. J.; Green, A. An international database for pesticide risk assessments and management. *Hum. Ecol. Risk Assess.* **2016**, *22* (4), 1050−1064.

(42) Anastas, P. T.; Warner, J. C. *Green Chemistry: Theory and Practice*; Oxford University Press, 1998.

(43) Yan, S. W.; Song, W. H. Photo-transformation of pharmaceutically active compounds in the aqueous environment: a review. *Environ. Sci. Proc. Impacts* **2014**, *16* (4), 697−720.

(44) Devi, L. G.; Rajashekhar, K. E. A kinetic model based on non-linear regression analysis is proposed for the degradation of phenol under UV/solar light using nitrogen doped TiO2. *J. Mol. Catal. a-Chem.* **2011**, *334* (1−2), 65−76.

(45) US EPA National Center for Computational Toxicology. Aggregated Computational Toxicology Resource Database (ACToR).

(46) Teather, K.; Parrott, J. Assessing the chemical sensitivity of freshwater fish commonly used in toxicological studies. *Water Qual. Res. J. Can.* **2006**, *41* (1), 100−105.

(47) Daniel, A. B.; Choksi, N.; Abedini, J.; Bell, S.; Ceger, P.; Cook, B.; Karmaus, A. L.; Rooney, J.; To, K. T.; Allen, D. Data curation to support toxicity assessments using the Integrated Chemical Environment. *Front. Toxicol.* **2022**, *4*, No. 987848.

(48) Kostal, J.; Plugge, H.; Raderman, W. Quantifying Uncertainty in Ecotoxicological Risk Assessment: MUST, a Modular Uncertainty Scoring Tool. *Environ. Sci. Technol.* **2020**, *54* (19), 12262−12270.

(49) Beck, N. B.; Becker, R. A.; Erraguntla, N.; Farland, W. H.; Grant, R. L.; Gray, G.; Kirman, C.; LaKind, J. S.; Lewis, R. J.; Nance, P. Approaches for describing and communicating overall uncertainty in toxicity characterizations: US Environmental Protection Agency's Integrated Risk Information System (IRIS) as a case study. *Environ. Int.* **2016**, *89−90*, 110−128.

(50) Plugge, H.; Das, N.; Kostal, J. Toward a Universal Acute Fish Threshold of Toxicological Concern. *Environ. Toxicol. Chem.* **2021**, *40* (6), 1740−1749.

(51) Przybylak, K. R.; Madden, J. C.; Cronin, M. T.; Hewitt, M. Assessing toxicological data quality: basic principles, existing schemes and current limitations. *SAR QSAR Environ. Res.* **2012**, *23* (5−6), 435−459.

(52) Russom, C. L.; Bradbury, S. P.; Broderius, S. J.; Hammermeister, D. E.; Drummond, R. A. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). *Environ. Toxicol. Chem.* **1997**, *16* (5), 948−967.

(53) Melnikov, F.; Kostal, J.; Voutchkova-Kostal, A.; Zimmerman, J. B.; Anastas, P. T. Assessment of predictive models for estimating the acute aquatic toxicity of organic chemicals. *Green Chem.* **2016**, *18*, 4432−4445.

(54) Ramer, S.; Kostal, J. An Accurate way of Predicting the Aquatic Toxicity of Pesticides. In *International Symposium on Green Chemistry*; La Rochelle, France, 2019.

(55) US EPA., 2012, 1−43, 2012, 1−43. http://www.epa.gov/sites/production/files/2015-09/documents/ecosartechfinal.pdf (accessed March 2015).

(56) US EPA, 2010. http://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test (accessed March 2015).