Environmental Toxicology and Chemistry, 2025, 1–14
https://doi.org/10.1093/etojnl/vgaf046
Advance access publication: February 12, 2025
Original Article

Special Series

Addressing the challenges of acute fish toxicity hazard classification using a non-animal defined approach

Donna S. Macmillan[1,*,‡], Pravin Ambure[2], Vicente Aranda[2], Yannick Bayona[3], Vera Bonderovic[4,5], Jay Dawick[6], Nicolas Fabre[3], Stephan Fischer[7], Geoff Hodges[8], Ágata Llobet-Mut[2], Sophie Loisel-Joubert[3], Claudia Rivetti[8], Jayne Roberts[8], Kristin Schirmer[9], Eva Serrano-Candelas[2], Blanca Serrano Ramón[10], and Ricky A. Stackhouse[11]

[1]Humane Society International, Washington, DC, United States
[2]ProtoQSAR, Paterna, Valencia, Spain
[3]L'Oréal Research and Innovation, Clichy, France
[4]REVIVO BioSystems Pte. Ltd, Singapore
[5]University of Geneva, Geneva, Switzerland
[6]Innospec Limited, Cheshire, United Kingdom
[7]aQuaTox-Solutions GmbH, Wallisellen, Switzerland
[8]Safety and Environmental Assurance Centre, Unilever, Sharnbrook, Bedfordshire, United Kingdom
[9]Department of Environmental Toxicology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
[10]ECETOC, Brussels, Belgium
[11]Sasol, Houston, TX, United States

*Corresponding author: Donna S. Macmillan. Email: dmacmillan@iccs-cosmetics.org

‡Present address: International Collaboration on Cosmetic Safety, NY, United States

Abstract

Acute fish toxicity is an ecotoxicological endpoint that provides important information about a chemical's potential to elicit (an) adverse effect(s) on fish. These effects are typically studied using in vivo tests, but for animal welfare reasons as well as the quest for increased species relevance, biological coverage, mechanistic understanding of effects, and throughput, there have been significant efforts in recent years to reduce or replace the use of animals in (eco)toxicological hazard assessment by developing defined approaches (DA) or integrated approaches to testing and assessment. To this end, a novel score-based DA has been developed as a proof-of-concept, which integrates three in silico predictions from freely available (quantitative) structure activity relationship models: the VEGA Fish (KNN-Read-Across) and Fathead Minnow (KNN-IRFMN) models and the United States Environmental Protection Agency ECOSAR Fish 96-h LC50 model, along with in vitro RTgill-W1 data. The DA provides a categorical output aligned with the United Nations Globally Harmonized System of Classification and Labelling framework (Acute Category 1, Acute Category 2, Acute Category 3, or Not Classified) with an overall accuracy of 80%, offering a reliable alternative to traditional in vivo testing methods for acute fish toxicity.

Keywords: aquatic acute toxicity, defined approaches, classification and labelling, Globally Harmonized System of Classification and Labelling

Introduction

The Animal-Free Safety Assessment Collaboration

Humane World for Animals (previously known as Humane Society International) leads a multi-stakeholder initiative known as the Animal-Free Safety Assessment (AFSA) Collaboration, which was developed to accelerate the global adoption of modern chemical safety assessment using non-animal approaches (Animal-Free Safety Assessment Collaboration, 2018). The AFSA Collaboration focuses on several key areas, including uniting leading industry and not-for-profit organizations with a shared goal of replacing animal testing with scientifically robust and protective non-animal approaches.

Hazard characterization and classification for acute aquatic toxicity

Acute aquatic toxicity is an ecotoxicological endpoint that provides important information about a chemical's potential to elicit (an) adverse effect(s) on aquatic organisms over short-term exposure(s). Within regulatory ecotoxicology, three trophic levels are typically considered as proxies of the ecosystem: fish (representing vertebrates/apical consumers), daphnia (representing invertebrates/primary consumers), and algae (representing plants/primary producers). Acute aquatic toxicity is typically studied using one or more Organisation for Economic Co-operation and Development (OECD) test guideline assays such as the Fish Acute Toxicity Test (OECD 203), the Fish Embryo Acute Toxicity Test

Received: October 1, 2024. **Revised:** January 31, 2025. **Accepted:** February 3, 2025

(FET; OECD 236), the Fish Cell Line Acute Toxicity—The RTgill-W1 cell line assay (OECD 249); Daphnia sp., Acute Immobilisation Test (OECD 202); Freshwater Alga and Cyanobacteria, Growth Inhibition Test (OECD 201); and the Freshwater aquatic plants, Lemna sp. Growth Inhibition Test (OECD 221; Organisation for Economic Co-Operation and Development, 2004, 2006, 2011, 2013, 2019, 2021b). The concentration that causes 50% lethal, immobilization, or growth inhibition effects, depending on the test guideline followed (i.e., the lethal or effect [L(E)C]50) values obtained from these assays) is used within several regulatory frameworks. These include performing environmental risk assessment of chemicals under the European Union (EU) regulation 1272/2008 on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH; European Union, 2006) or of plant protection products under EU regulation 1107/2009 concerning the placing of plant protection products on the market (European Union, 2009). They are also used within classification and labeling frameworks such as the Globally Harmonized System of Classification and Labelling of Chemicals (GHS) or the European Union Classification and Labelling and Packaging (EU CLP) regulation to assign categories of Acute 1, Acute 2, Acute 3, and Not Classified, which indicate the inherent degree of hazard of a given substance (Table 1; European Union, 2008; United Nations Globally Harmonized System, 2023).

Use of animal models for acute fish toxicity

In 2019, EU regulation 2019/1010 mandated that all Member States collect and publicly disclose annual statistics on animal use, which can be used to monitor the impact of 3Rs (Replace, Reduce, Refine animal use) initiatives (European Union, 2019). The 3Rs principle aims to avoid unnecessary animal testing and promote alternative approaches, whenever possible (Russell & Burch, 1960). In recent years, there has been a concerted effort to reduce or eliminate the use of vertebrate fish and increase the adoption of non-animal New Approach Methodologies (NAMs) for regulatory environmental hazard and risk assessments, driven by concerns for animal welfare and the desire for greater species relevance (through NAMs designed to evaluate a common mechanism), biological coverage, and improved throughput.

Depending on their life stage (e.g., excluding larval stages), fish are classified as animals under the EU definition as stated in Directive 2010/63/EU and its amendments (European Union, 2010). Despite reports of a decline in short-term fish toxicity studies (European Chemicals Agency, 2023), the ALURES database (European Commission, 2022), which allows filtering by use, shows that approximately 30,000 fish were specifically used for regulatory testing in 2022.

Combining information sources to increase relevance, protection, and predictivity power

The OECD has published several frameworks to assist regulatory decision-making on various specific topics (Organisation for Economic Co-operation and Development, 2017a, 2017b, 2019, 2023, 2024). These approaches include integrated approaches to testing and assessment (IATA) and defined approaches (DA). The former, IATA, requires expert judgment, whereas DAs use a defined set of information sources in a fixed data interpretation procedure (e.g., decision trees or complex algorithms) to address a specific regulatory need. The information sources used in DAs may incorporate a range of complementary methods, including in silico approaches ([quantitative] structure activity relationship models; [(Q)SARs]), mechanistic models (physiologically based pharmacokinetic, toxicokinetics/toxicodynamics), as well as in chemico and in vitro methods, collectively known as NAMs. Several DAs, such as those for skin sensitization and eye irritation, have been published as OECD guidelines (Organisation for Economic Co-Operation and Development, 2023, 2024), with at least one DA incorporating (Q)SAR predictions. There is growing interest in expanding their chemical and toxicological applicability and their use in various contexts (Alépée et al., 2019, 2023; Alépée & Adriaens, 2024; Organisation for Economic Co-Operation and Development, 2023, 2024; Strickland et al., 2022).

In silico (Q)SAR models have been available to predict acute aquatic toxicity for many years (Cronin, 2017). The (Q)SARs are relevant within certain domains of applicability and take into account many properties, including chemical class and structure, mode/mechanism of action (MoA), physical and chemical properties, and model organisms, leading to a wide choice of available (Q)SAR models and versions. (Quantitative) structure activity relationship models are permitted for use in a regulatory context providing they fulfill the (Q)SAR validation principles (Organisation for Economic Co-operation and Development, 2014), and they are often used in weight of evidence approaches alongside other information sources such as in vitro assays (Voigt & Jaeger, 2023). Regulatory agencies such as the European Chemicals Agency (ECHA) have also published guidance on the use and reporting of (Q)SARs (European Chemicals Agency, 2008, 2016).

Consequently, this project aimed to prioritize the use of open-source and freely available (Q)SAR models and the in vitro RTgill-W1 assay to develop a DA using a dataset compiled of publicly available data. The main objective was to create a DA that is accessible to a broad audience of experts and capable of predicting acute fish toxicity classifications suitable for EU CLP and/or GHS classification contexts. The FET was not considered as a line of evidence/potential information source in the DA because this assay is not considered to be a non-animal study by all stakeholders/organizations.

Table 1. An overview of the median lethal or effect concentration values and their corresponding hazard classification, as defined by European Union Classification and Labelling and Packaging (EU CLP) or Globally Harmonized System (GHS) of Classification and Labelling of Chemicals. NC = not classified.

Concentration (mg/L)	≤ 1	>1 ≤ 10	>10 ≤ 100	>100
Classification	Acute 1	Acute 2	Acute 3	NC
Applicable under EU CLP?	Yes	No	No	Yes
Applicable under GHS?	Yes	Yes	Yes	Yes

Materials and methods

Compilation of a dataset used to develop the defined approach (AFSA dataset)

Several data sources were integrated using the Konstanz Information Miner (KNIME Analytics Platform Ver. 5.2.4; Berthold et al., 2009) to create a comprehensive dataset (herein referred to as the "AFSA dataset") for use to develop the DA, including several literature sources and the ECOTOX database as extracted from OECD QSAR Toolbox (Lillicrap et al., 2020; Natsch et al., 2018; Organisation for Economic Co-Operation and Development, 2021a; Tanneberger et al., 2013). Each data source was standardized and curated by unifying chemical names, renaming column headers, filtering out RTgill-W1 studies reporting nominal values (retaining measured values only), making units consistent, removing non-test guideline species and excluding entries with non-test guideline study durations (retaining 96 hr studies for

acute fish toxicity; 24 hr studies for RTgill-W1 studies), removing nonstandard numerical values (e.g., those using qualifies such as >, <), and, finally, by removing superfluous columns. The data sources were then aggregated by Chemical Abstracts Service Registry Number (CAS RN). During the integration process, it was apparent that several data sources contained duplicate results from the same studies. Therefore, to reduce "double-counting" of the same experimental value (because manual curation was not possible due to size of dataset), a ranking system was applied to the data sources. If ECOTOX data were available for a specific substance, this was preferentially used, due to its extensive coverage from various sources. The remaining publications were prioritized in the following order: Lillicrap et al. 2020, Tanneberger et al., 2013, and, finally, Natsch et al. (2018), to obtain additional data for specific substances not previously identified. After applying the ranking, if more than one study result was available for a unique substance in the data source, the geometric mean value per species, as well as across all species, was calculated (Figure 1). The ranking system was not intended to reflect the quality of individual data sources but was used to ensure that the experimental data in the AFSA dataset were as reliable as possible, to create a quality dataset for use in the development of the DA. The final step was to filter the dataset to keep only those chemicals that also had a data point from the RTgill-W1 assay. Extensive data curation was outside the scope of the project.

Simplified Molecular-Input Line-Entry System (SMILES) are a compact way of representing a chemical substance's structure (International Union of Pure and Applied Chemistry, 1997b), and SMILES strings are often used as input into (Q)SARs. For substances where SMILES was not reported in the data source, ProtoQSAR's in-house Python script was used to automatically retrieve SMILES from PubChem using the CAS RN (Kim et al., 2023; National Institutes of Health, 2024b; Python Software Foundation, 2024). An additional Python script, developed by ProtoQSAR, using functions from the rdkit package (RDKit: Open-source cheminformatics, 2021), was used to curate the SMILES; this includes removing incorrect molecules, inorganic and organometallic molecules from the dataset, sanitizing the SMILES string molecular structures (by harmonizing atom numbering, checking valencies, setting aromaticity, conjugation and hybridization, etc.), eliminating the counterions of organic salts, and identifying and combining duplicate molecules. A manual curation step was performed to remove mixtures.

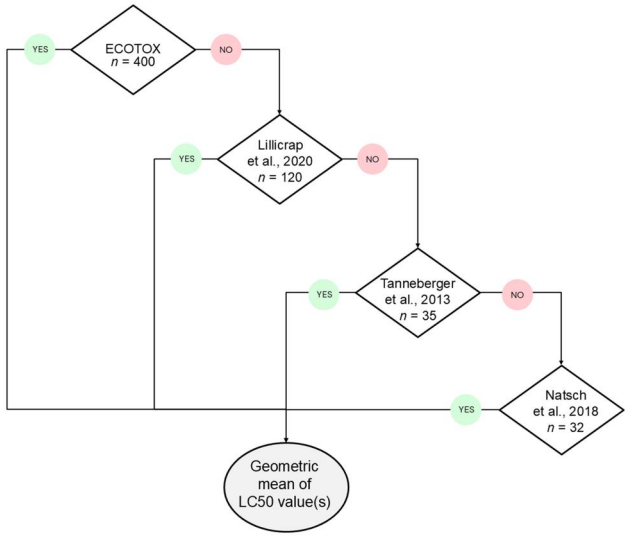


Figure 1. Workflow depicting ranking of data sources used to derive fish median lethal concentration (LC50) values in the Animal-Free Safety Assessment (AFSA) dataset. If fish LC50 data was present in ECOTOX, this was used preferentially. The remaining publications were then searched for fish LC50 data in the order of Lillicrap et al. (2020), Tanneberger et al. (2013), and, finally, Natsch et al. (2018), to obtain additional data for specific substances not previously identified. The ranking is not a reflection of the perceived quality of data contained in the individual data sources.

The final AFSA dataset consisted of 405 unique substances with fish LC50 data, and it was used to evaluate, and then select, the (Q)SAR models to be used in the DA. A subset of 66 chemicals also had corresponding RTgill-W1 data, which was used to evaluate the DA. The final composition of the dataset is illustrated in Table 2. The AFSA dataset is provided as online supplementary material.

Chemical space analysis

To evaluate the coverage of the AFSA dataset, a chemical space analysis was conducted. Several public access databases were collected, representing diverse molecular chemical space. The first was the ZINC 15 database (Sterling & Irwin, 2015), a collection of commercially available "drug-like" compounds, obtained from rdkit as mol format, a format used to encode chemical structures as text-based connection tables (International Union of Pure and Applied Chemistry, 1997a), and SMILES notation. Next, a large set of fragrance ingredients was acquired from the International Fragrance Association Transparency List as CAS RN (International Fragrance Association, 2024). The SMILES data were retrieved from the PubChem database using the CAS RN, and from the CACTVS server if the SMILES were not retrievable by PubChem, using automatic Python scripts (National Institutes of Health, 2024a, 2024b). The SMILES of substances corresponding to industrial products were retrieved from the ECHA REACH database, which contains registered chemical substances provided by companies in accordance with the requirements of the REACH Regulation (European Union, 2006), and from the Toxic Substances Control Act (TSCA) database (United States Environmental Protection Agency, 2013), which contains all existing chemical substances manufactured, processed, or imported in the United States, by means of QSAR Toolbox (Organisation for Economic Co-Operation and Development, 2021a). A database of 437,859 natural products was also compiled. The data to construct this database was extracted from different online sources containing information about natural products and nutraceuticals: 406,747 molecules from the COlleCtion of Open NatUral producTs (COCONUT) database (Sorokina et al., 2021), 107 nutraceuticals from Drugbank (Wishart et al., 2018), 20 nutraceuticals from Drugs.com (Drugs.com, 2024), 59 nutraceuticals from FooDB (FooDB, 2024), and 30,926 natural products from the Natural Product Activity and Species Source (NPASS) database (Zeng et al., 2018). A curating

Table 2. Composition of the Animal-Free Safety Assessment (AFSA) dataset used to develop the defined approach for acute fish toxicity.

Taxa	Species	Data sources	*n*
Fish	*Cyprinus carpio, Danio rerio, Lepomis macrochirus, Oryzias latipes, Oncorhynchus mykiss, Pimephales promelas*	ECOTOX, Natsch et al. (2018), Lillicrap et al. (2020), Tanneberger et al. (2013)	405
RTgill-W1	*O. mykiss*	Natsch et al. (2018), Tanneberger et al. (2013)	66

process was performed on the initial database to remove duplicated molecules, salts, inorganic compounds, and mixtures, obtaining a resulting database of 429,460 compounds. The natural products datasets were then merged and reduced to 10% of its original size by clustering it using the density-based Ordering Points To Identify the Clustering Structure (OPTICS) method, with the Jaccard index, based on Molecular ACCess Systems (MACCS) fingerprints (Ankerst et al., 1999). This reduction method results in an easier-to-handle dataset (~40k chemicals compared to the original ~400k) with the same approximate representation of chemical groups.

Chemical structures from these different databases as well as from the AFSA dataset were encoded as Morgan fingerprints to represent the chemical entities and the spatial connectivity in each substance, for comparative analysis (Morgan, 1965). Chemical space was compared among them using the two most significant principal components derived from the principal component analysis (PCA; Greenacre et al., 2022).

The databases were also analyzed using physicochemical properties: boiling point (BP), octanol/water distribution ratio (logD), octanol/water partition coefficient (logP, often referred to as logKow), melting point (MP), and water solubility (WS). The physicochemical properties were calculated using the ProtoPHYSCHEM module, which is a compendium of QSAR models that predict physicochemical properties integrated in web tool ProtoPRED (ProtoQSAR, 2025).

(Q)SAR models

Several software tools containing (Q)SAR models for the evaluation of acute fish toxicity were collected, with the aim of evaluating them against the AFSA dataset. This evaluation aimed to select appropriate (Q)SAR models for use in the DA. When collecting possible software tools, only noncommercial, open-access (freely available) software was considered. Moreover, additional aspects were considered during the software selection: their ease of use, level of acceptance within the experimental field, and their capacity to perform batch predictions for large datasets. Three software tools consisting of 12 (Q)SARs were evaluated, VEGA, United States Environmental Protection Agency (USEPA) ECOSAR, and T.E.S.T. These (Q)SARs are, of course, not an exhaustive list of those that can be used for this toxicological endpoint, and there may be other (Q)SARs that are also applicable for use in a DA, e.g., iSafeRat (KREATiS, 2024).

VEGA

The VEGA platform provides a series of freely available (Q)SAR models for regulatory purposes, and the following eight models were used in this project: Fish Acute (LC50) Toxicity model (IRFMN), Fish Acute (LC50) Toxicity model (NIC), Fish Acute (LC50) Toxicity model (KNN-Read-Across), Fish Acute (LC50) Toxicity classification (SarPy-IRFMN), Fish Acute (LC50) Toxicity model (IRFMN-Combase), Fathead Minnow LC50 96 h (USEPA), Fathead Minnow LC50 model (KNN-IRFMN), and Guppy LC50 model (KNN-IRFMN). VEGA Ver. 1.2.3 (Benfenati et al., 2013) was used and predictions from its eight models generated using SMILES and the batch prediction mode. The VEGA platform provides predicted LC50 values (except for Fish Acute (LC50) Toxicity classification (SarPy-IRFMN), which provides classes corresponding to GHS categories), and an estimation of the reliability of the prediction. The reliability of the prediction is related to the applicability domain (AD) of the model as defined by the developers (Danieli et al., 2023). An AD index is provided with each prediction: if the AD index is between 1 and 0.85, the substance is regarded within the AD of the model and corresponds to good reliability. If the AD index is between 0.85 and 0.7, the substance could be out of the AD of the model and corresponds to moderate reliability. If the AD index is less than 0.7, the substance is regarded out of the AD of the model and corresponds to a low reliability prediction. Only predictions with good or moderate reliability were considered. The predicted LC50s from the eight VEGA models (except the Fish Acute (LC50) Toxicity classification (SarPy-IRFMN) model, which directly provided predictions as GHS classifications; Table 1) were converted to the respective acute environmental GHS classifications and corresponding DA scores for additional analyses using KNIME.

USEPA ECOSAR

The Ecological Structure Activity Relationships (ECOSAR) Predictive Model, developed by the USEPA, is a freely available model that predicts aquatic toxicity based on the applicable/assigned chemical class of the substance under assessment (Reuschenbach et al., 2008). Two models were used, one for freshwater fish, ECOSAR Fish 96-h LC50, and one for saltwater fish, ECOSAR Fish 96-h LC50 (SW). The ECOSAR model Ver. 2.2 (United States Environmental Protection Agency, 2024) was used to generate predictions from both models using the SMILES string for each substance. For substances with more than one prediction (due to classification into multiple chemical classes, e.g., organic neutral, amides, vinyl, esters), the most conservative value was used according to the precautionary principle. Substances with a Log Kow above the Max Log Kow for each ECOSAR class, as defined by ECOSAR, were considered as outside the AD and excluded from the analysis. The predicted LC50s from each ECOSAR model were converted to the respective acute environmental GHS classifications and corresponding DA scores for additional analyses using KNIME.

T.E.S.T

The Toxicity Estimation Software Tool is a freely available tool from the USEPA and consists of several methodologies, which are grouped as two (Q)SARs, the single model and the consensus model. The single model method (a multilinear regression model) and the consensus model (an average of the predictions from all (Q)SAR methodologies in T.E.S.T.; single-method, hierarchical, group contribution, nearest neighbor and mode of action methods) was used. The T.E.S.T. tool Ver. 5.1.2.0 was used to generate predictions from its single method model and consensus model using the SMILES string for each substance (Martin et al., 2023). When working in single model mode, and the software does not provide a prediction, this means that the molecule does not fall within the AD. When using the consensus model, which is the average of the predicted toxicities from all (Q)SAR methodologies, the applicability domain of each method is taken into account, and if only one methodology provides a prediction, then the prediction was not used. Experimental values retrieved by T.E.S.T. were not considered. The predicted LC50s from T.E.S.T. models were converted to the respective acute environmental GHS classifications and corresponding DA scores for additional analyses using KNIME.

Abbreviations of models

The models were abbreviated as follows: VEGA Fish Acute (LC50) Toxicity model (IRFMN) = Fish (IRFMN); VEGA Fish Acute (LC50) Toxicity model (NIC) = Fish (NIC); VEGA Fish Acute (LC50) Toxicity model (KNN-Read-Across) = Fish (KNN-Read-Across); VEGA Fish Acute (LC50) Toxicity classification (SarPy-IRFMN) = Fish (SarPy-IRFMN); VEGA Fish Acute (LC50) Toxicity model (IRFMN-Combase) = Fish (IRFMN-Combase); VEGA Fathead

Minnow LC50 96 h (USEPA) = Fathead Minnow (USEPA); VEGA Fathead Minnow LC50 model (KNN-IRFMN) = Fathead Minnow (KNN-IRFMN); VEGA Guppy LC50 model (KNN-IRFMN) = Guppy (KNN-IRFMN); ECOSAR Fish 96-h = LC50 ECOSAR; ECOSAR Fish 96-h LC50 (SW) = ECOSAR (SW); T.E.S.T. (single model) = T.E.S.T (single); T.E.S.T. (consensus) = T.E.S.T. (consensus).

MoA profiler (MechoA)

The mode of action of the substances in the AFSA dataset was assessed using the MechoA profiler (Ver. 1.0) available in the OECD QSAR Toolbox (Organisation for Economic Co-Operation and Development, 2021a) using the SMILES input. The MechoA profiler classifies substances into six general MechoAs and 23 MechoA subgroups (Bauer et al., 2018). To simplify the evaluation of the DA against discrete MoA, each output from the individual programs was grouped into the broad categories of polar narcotics, nonpolar narcotics, reactive (unspecific), reactive (specific), and so on. See online supplementary material for assignment of MechoA classifications into the broad categories.

Metrics

Several metrics are used throughout the article to describe the performance of the individual (Q)SARs and the DA. In all instances, the experimental result refers to the fish LC50 data (converted to GHS) from the AFSA dataset, and the (Q)SAR/DA prediction refers to the GHS prediction from the (Q)SAR or the DA.

For example, if the DA or (Q)SAR predicts Not Classified and the fish GHS classification is Acute 1, then this would be considered a false negative prediction. If the DA or (Q)SAR predicts Acute 1 and the fish GHS classification corresponds to Acute 2, then this would be considered to be a false positive prediction.

True positive (TP) = Positive experimental result and positive (Q)SAR/DA prediction; false positive (FP) = Negative experimental result and positive (Q)SAR/DA prediction; true negative (TN) = Negative experimental result and negative (Q)SAR/DA prediction; false negative (FN) = Positive experimental result and negative (Q)SAR/DA prediction; sensitivity (Se) = TP/[TP + FN]; specificity (Sp) = FP/[FP + TN]; accuracy (Acc) = [TP = TN]/[TP + FP + TN + TN]; balanced accuracy (BA) = Se + Sp/2.

Results

Chemical space analysis

To evaluate the chemical space covered by the AFSA dataset, it was compared against the chemical space of diverse datasets consisting of industrial chemicals (Organisation for Economic Co-Operation and Development, 2021a), drugs (ZINC 15, 2024), natural products (Drugs.com, 2024; FooDB, 2024; Sorokina et al., 2021; Wishart et al., 2018; Zeng et al., 2018), and fragrances (International Fragrance Association, 2024). A PCA was performed using the two most significant principal components (Figure 2). Principle component analysis is a data visualization technique that reduces the dimensionality of a dataset to its essential features, without losing information. This increases the interpretability of a given dataset, because the principal components are used as variables, which facilitates direct comparison of the variability across multiple datasets (Greenacre et al., 2022).

By comparing the AFSA dataset (dataset_AFSA) against the other datasets, we can conclude that our dataset represents enough heterogeneity to cover substances of a diverse nature, as illustrated by the overlapping of the chemical spaces in the PCA. For example, there is good coverage of the AFSA dataset across industrial chemicals (left panel: dataset_TSCA, dataset_REACH) as well as drugs, natural products, and fragrances (right panel: dataset_Drugs, dataset_NatProd, Dataset_Fragrances, respectively), increasing confidence that the dataset used to develop the DA represents a wide and diverse chemical space; thus, the DA described can be applied to substances across various chemical sectors.

The datasets described above (and the RTgill-W1 dataset) were also analyzed by comparing the following physicochemical properties: BP, logD, logP, (often referred to as logKow), MP, and WS. The resulting violin plots, reported in online supplementary material S1, supported the PCA analysis and indicated that the dataset used to develop the DA consisted of a diverse set of chemicals. In addition, the chemical space was also visualized using correct and incorrect DA predictions, and this is reported in online supplementary material S2.

Predictivity of (Q)SARs

Each (Q)SAR model was evaluated against the acute fish LC50s in the AFSA dataset, after converting each (Q)SAR prediction and

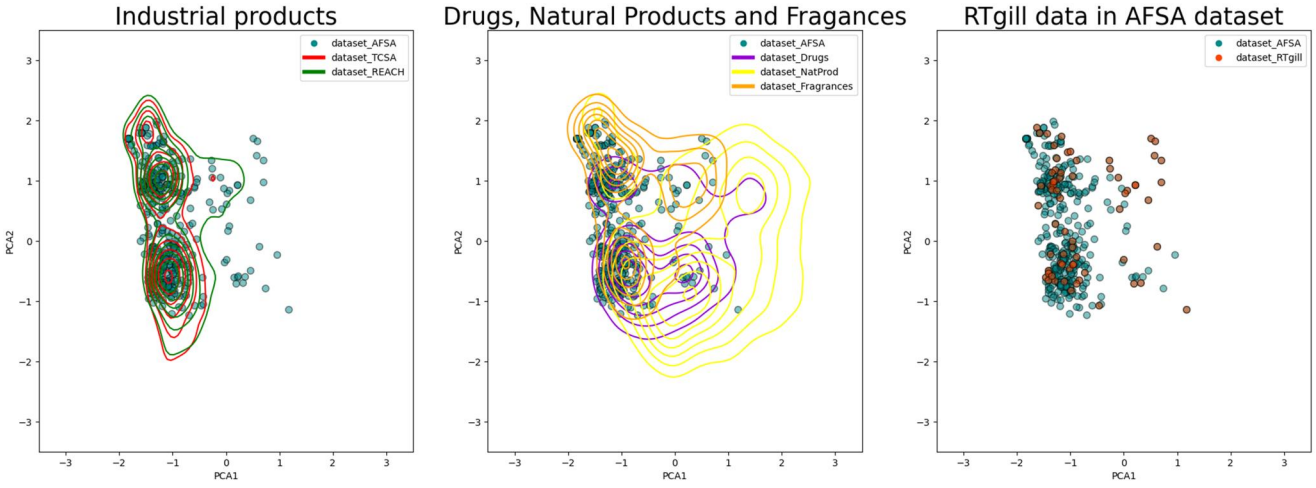
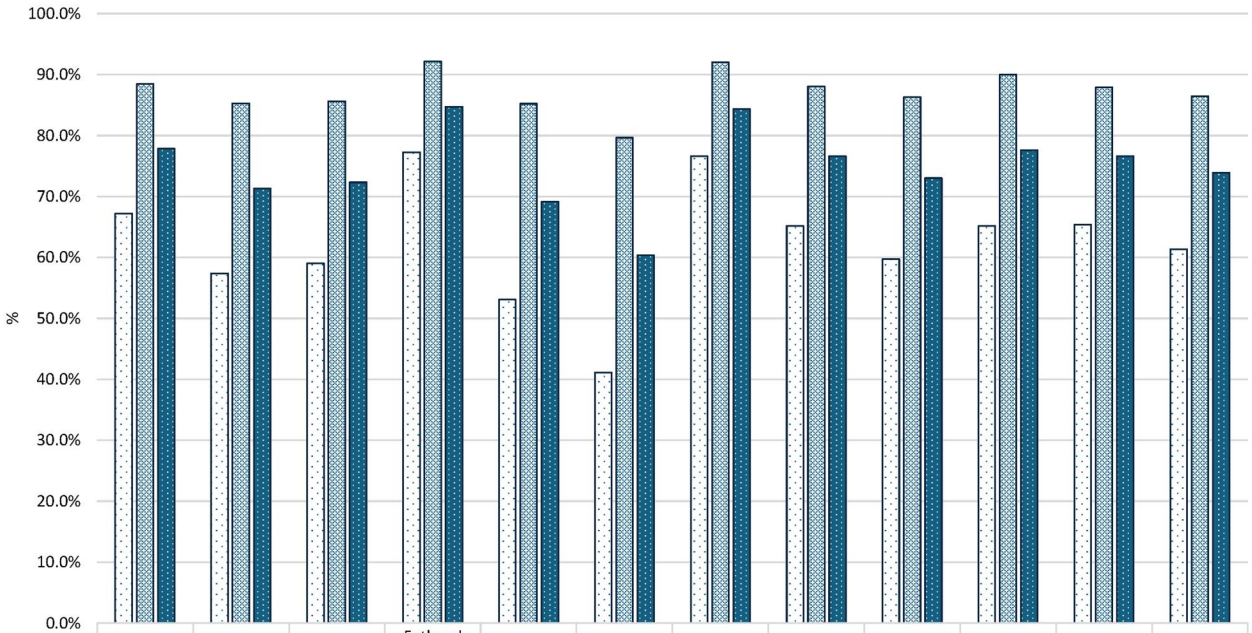


Figure 2. Chemical space comparison using principal component analysis (PCA). Left panel: The distributions of Principal Component 1 (PC1) and Principal Component 1 (PC2) between industrial chemicals (dataset_REACH and dataset_TCSA) and the Animal-Free Safety Assessment (AFSA) dataset (dataset_AFSA; dots). Middle panel: The distributions of PC1 and PC2 between U. S. Food and Drug Administration-approved drugs (dataset_Drugs), fragrances (dataset_Fragrances) and natural products (dataset_NatProd) and the AFSA dataset (dataset_AFSA; dots). Right panel: Representation of the chemical space of the AFSA dataset (dataset_AFSA; light dots) with those chemicals that have RTgill-W1 data highlighted as darker dots.



	ECOSAR	ECOSAR (SW)	Fathead Minnow (EPA)	Fathead Minnow (KNN-IRFMN)	Fish (IRFMN)	Fish (IRFMN-Combase)	Fish (KNN-Read-Across)	Fish (NIC)	Fish (SarPy-IRFMN)	Guppy (KNN-IRFMN)	T.E.S.T (consensus)	T.E.S.T (single)
Sens	67.2%	57.4%	59.0%	77.2%	53.1%	41.1%	76.6%	65.2%	59.7%	65.2%	65.4%	61.3%
Spec	88.5%	85.3%	85.6%	92.1%	85.2%	79.6%	92.0%	88.0%	86.3%	90.0%	87.9%	86.5%
BA	77.8%	71.3%	72.3%	84.7%	69.1%	60.3%	84.3%	76.6%	73.0%	77.6%	76.6%	73.9%

Figure 3. Performance metrics of individual (quantitative) structure-activity relationship ((Q)SAR) models ((Q)SARs) evaluated against the fish median lethal concentrations (converted to Globally Harmonized System of Classification and Labelling of Chemicals categories) in the Animal-Free Safety Assessment dataset. *Note*. BA = balanced accuracy, Sens = sensitivity, Spec = specificity.

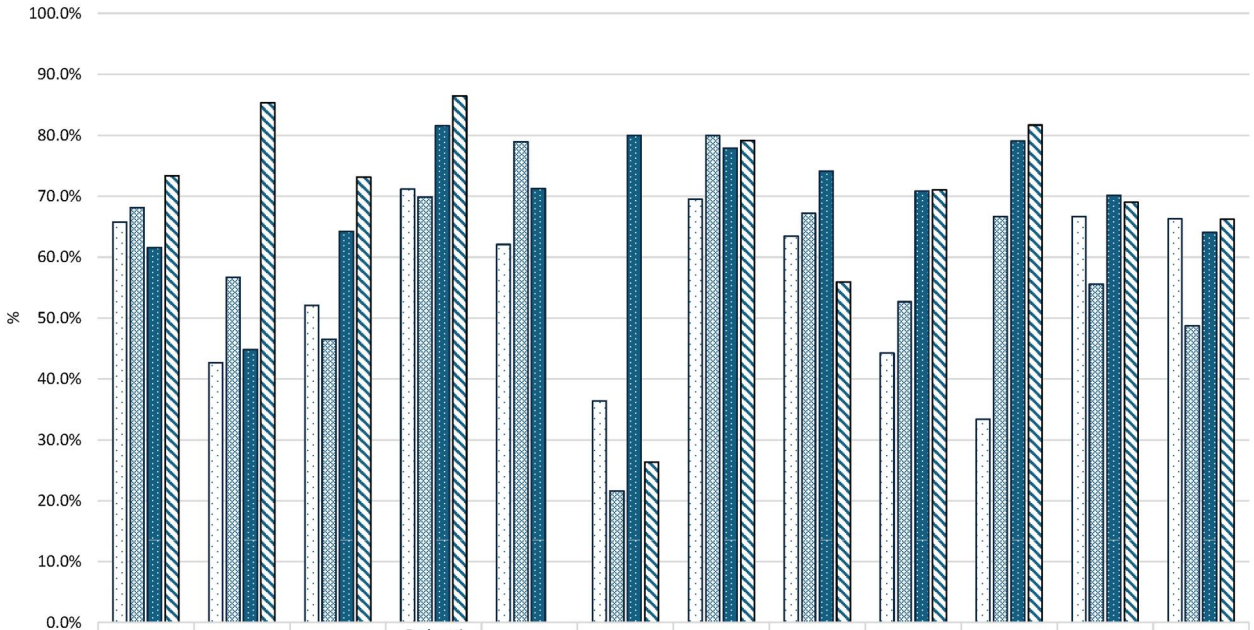
acute LC50 to the corresponding GHS classification. Where possible, low reliability (Q)SAR predictions were removed before the calculation of performance metrics. The performance metrics of each (Q)SAR were evaluated against a maximal subset of the dataset (i.e., all substances with a prediction from any individual (Q)SAR; *n* ranging between 113 and 383 depending on number of predictions possible from the (Q)SAR under evaluation). After evaluation, three models were selected for use in the DA, based on those with the highest overall balanced accuracy and by-class sensitivity: VEGA Fish (KNN-Read-Across) and VEGA Fathead Minnow (KNN-IRFMN) model and the USEPA ECOSAR Fish 96-h LC50 model, with BAs of 84.3%, 84.7%, and 77.8%, respectively (Figures 3 and 4).

Development and performance of score-based DA

When considering how to develop a DA to predict the GHS classification of acute fish toxicity, it was decided to use a score-based approach, similar to the Integrated Testing Strategy (ITS) defined approach for skin sensitization, published as an OECD guideline in 2021 and later updated in 2023 (Organisation for Economic Co-Operation and Development, 2023). Filtering the AFSA dataset to only include chemicals with a result from the RTgill-W1 assay, and a prediction from each (Q)SAR model, the scoring was applied as follows: a score of 4 was given to (Q)SAR predictions and RTgill-W1 outputs corresponding to a classification of Acute 1 under GHS. The L(E)C50 values corresponding to a classification of Acute 2 and 3 were assigned scores of 2 and 1, respectively, and a (Q)SAR or RTgill-W1 result of Not Classified prompted a score of 0. A higher score of 4 for Acute 1 was chosen in order to provide a more conservative DA prediction, as desired (less likely to lead to false negatives). The mean score from the 3 (Q)SAR predictions and RTgill-W1 output was calculated and rounded following standard mathematical rounding rules and used to assign a GHS classification as shown in the DA (Figure 5). For example, if the mean score was over 3, then it was assigned as Acute 1, if the mean score was 2, then it was assigned as Acute 2, the mean score equaling 1 led to assignment as Acute 3, and a mean score of 0 led to an assignment of Not Classified.

A minimum of two (Q)SAR predictions, from the three permitted for use in the DA, and corresponding RTgill-W1 result for a substance must be available in order to apply the DA, and low reliability or out of domain predictions (as described by the individual (Q)SARs, where possible) should not be used. The DA must include (Q)SAR predictions and an experimental value from the RTgill-W1 assay to ensure diversity in the information sources used in the DA. If the DA was composed of similar information sources, that is, only (Q)SAR predictions, there is a possibility that the DA would be less likely to predict well because similar training sets may have been used in their development.

Overall, the DA predicts Acute 1, Acute 2, Acute 3, and Not Classified categories for acute fish toxicity with an overall accuracy of 80%. The performance by-class (Acute 1, Acute 2, Acute 3, and Not Classified) ranges from 73.3% for Acute 2 and Acute 3 to 90.9% for Acute 1, and 87.5% for Not Classified (Table 3). The DA can also be used when only two (Q)SAR predictions are available, to allow for instances where one (Q)SAR may provide an out of domain prediction. The accuracy of the DA when either two or three (Q)SAR predictions are used decreases slightly to 78%. The performance metrics for this approach are reported in online supplementary material S3.



	ECOSAR	ECOSAR (SW)	Fathead Minnow (EPA)	Fathead Minnow (KNN-IRFMN)	Fish (IRFMN)	Fish (IRFMN-Combase)	Fish (KNN-Read-Across)	Fish (NIC)	Fish (SarPy-IRFMN)	Guppy (KNN-IRFMN)	T.E.S.T. (consensus)	T.E.S.T (single)
Acute 1	65.7%	42.6%	52.1%	71.1%	62.1%	36.4%	69.5%	63.4%	44.2%	33.3%	66.7%	66.3%
Acute 2	68.1%	56.6%	46.5%	69.9%	78.9%	21.6%	80.0%	67.2%	52.6%	66.7%	55.6%	48.7%
Acute 3	61.5%	44.8%	64.2%	81.5%	71.2%	80.0%	77.9%	74.1%	70.9%	79.0%	70.2%	64.0%
NC	73.3%	85.3%	73.1%	86.4%	0.0%	26.3%	79.1%	55.9%	71.0%	81.6%	69.0%	66.2%

Figure 4. By-class sensitivity of individual (quantitative) structure-activity relationship ((Q)SAR) models ((Q)SARs) evaluated against the fish median lethal concentrations (converted to Globally Harmonized System of Classification and Labelling of Chemicals categories) in the Animal-Free Safety Assessment dataset. *Note.* BA = balanced accuracy, Sens = sensitivity, Spec = specificity.

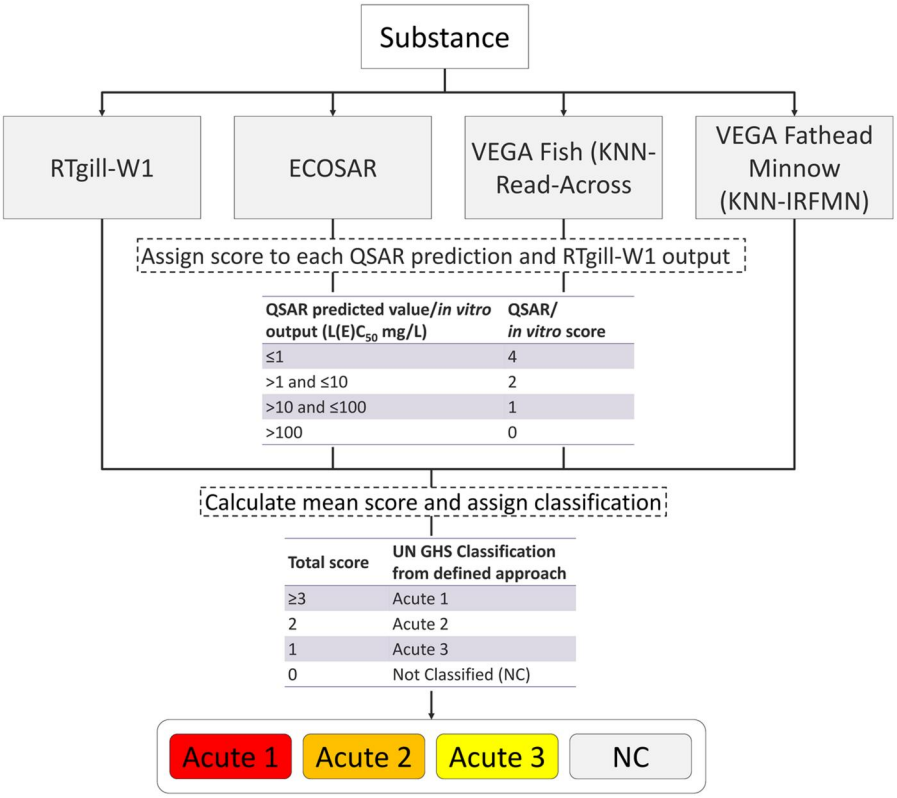


Figure 5. Score-based defined approach for acute fish toxicity using the RTgill-W1 assay and three (quantitative) structure-activity relationship ((Q)SAR) models, ECOSAR Fish 96-h median lethal concentration, VEGA Fish (KNN-Read-Across), and VEGA Fathead Minnow (KNN-IRFMN).

Table 3. Performance metrics of acute fish toxicity defined approach (DA).

		Predicted Globally Harmonized System (GHS) classification based on score-based DA			
		Acute 1	**Acute 2**	**Acute 3**	**Not Classified**
GHS classification based on experimental fish LC50	Acute 1	10	0	1	0
	Acute 2	2	11	2	0
	Acute 3	0	2	11	2
	NC	0	0	1	7
	Acc = 80%, $n = 49$				

Statistics of correct predictions and over/under predictions per GHS classification

Classification	**Acute 1**	**Acute 2**	**Acute 3**	**NC**
Correct %	90.9	73.3	73.3	87.5
Underpredicted % (false negative)	9.1 (Acute 3)	13.3 (Acute 3)	13.3 (NC)	–
Overpredicted % (false positive)	–	13.3 (Acute 1)	13.3 (Acute 2)	12.5 (Acute 3)
n	11	15	15	8

Note. LC50 = median lethal concentration; NC = not classified.

Discussion

Impact of using (Q)SARs in addition to RTgill-W1 assay

The in vitro RTgill-W1 assay has been suggested for use as a replacement for the in vivo acute fish toxicity test because it accurately predicts fish acute toxicity for a diverse range of chemicals (Natsch et al., 2018; Organisation for Economic Co-Operation and Development, 2021b; Tanneberger et al., 2013). This test is often used alongside multiple other information sources such as (Q)SAR predictions, physicochemical properties, and other lines of evidence. Consequently, the interest in developing DAs and IATAs has increased significantly in recent years, primarily for human health endpoints (Organisation for Economic Co-Operation and Development, 2023, 2024), but approaches for environmental endpoints have also been published (Lee et al., 2024). An analysis was undertaken to better understand if the (Q)SAR predictions were having a positive or negative effect compared to using the RTgill-W1 assay alone. For the subset of data with RTgill-W1 assay results, the assay alone predicts the correct GHS classification with 43% accuracy, whereas the DA predicts the same dataset with 80% accuracy ($n = 49$).

The relatively low accuracy of the RTgill-W1 can be attributed to two main issues. The first is misclassification of chemicals outside the applicability domain of the RTgill-W1 assay. For example, allyl alcohol is described in the corresponding OECD test guideline as a chemical that is not predicted well by the assay, presumably due to insufficient biotransformation to the toxic product, acrolein, which itself is well predicted by the RTgill-W1 assay. In the AFSA dataset, allyl alcohol is classified as Acute 1 using acute fish toxicity LC50 data, whereas the RTgill-W1 result corresponds to Not Classified. This example highlights the benefit of a DA, where other information sources can reduce the impact of an incorrect result/prediction. As two of the three (Q)SAR models predict allyl alcohol to be Acute 1 (ECOSAR predicting it to be Acute 3), the overall DA prediction is correct, and predicts it to be an Acute 1 fish toxicant. Another example is malathion, where the RTgill-W1 result corresponds to Acute 3, and the acute fish toxicity LC50 data indicates Acute 1. One known limitation of the RTgill-W1 assay is prediction of neurotoxins such as malathion, which inhibits acetylcholinesterase. Additionally, the oxidative metabolite malaoxon is more toxic than malathion, and is also a neurotoxin (Reed & Rubin, 2014). This substance has been discussed (Tanneberger et al., 2013), and analysis of the fish LC50 data (Fischer et al., 2019) highlighted a species sensitivity difference for this chemical, whereby the RTgill-W1 assay predicts malathion more in line with the less sensitive fish species (*Pimephales promelas*). The LC50s for malathion in the AFSA dataset range between 0.003 mg/L and 25 mg/L, across all species; therefore, applying the ranked geometric mean (for comparison of DA predictions) provides an LC50 value of 0.27 mg/L. If the LC50s of a single species of fish were considered, then the ranked geometric mean ranges from 0.077 mg/L for *Oncorhynchus mykiss* to 13.6 mg/L for *Pimephales promelas*, similar to the analysis reported by Fischer et al.

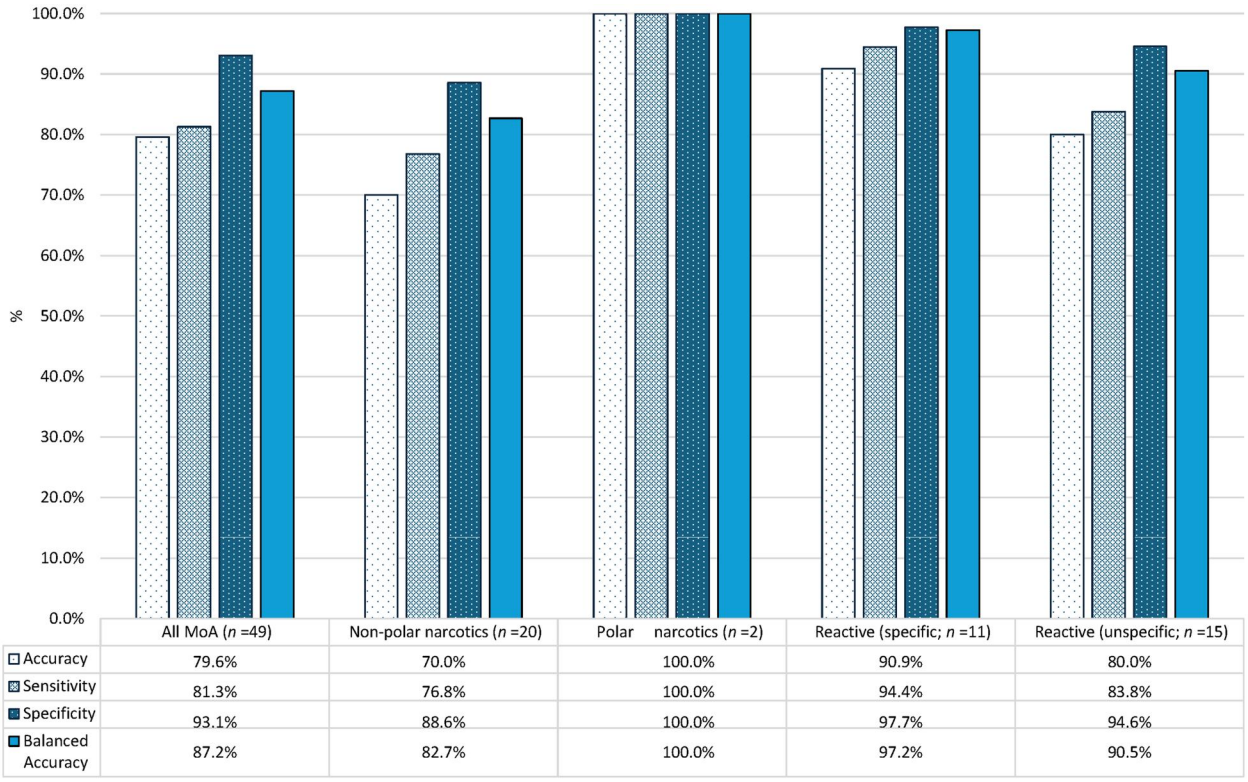
The second issue is the inherent difficulty of a classification system to distinguish between moderate or mild toxicity effects compared with strong and no effect, i.e., Acute 2/Acute 3 versus Acute/Not Classified in this case. This is reflected in this dataset, because the RTgill-W1 assays predicts 7/8 Not Classified chemicals, 2/15 Acute 3 chemicals, 6/15 Acute 2 chemicals, and 6/11 Acute 1 chemicals correctly. These results should be considered with caution, because this dataset is small and broad generalizations based on this analysis would be inappropriate.

For this dataset, in cases where the RTgill-W1 assay misclassifies, it tends to do this by only one GHS class, with the exception of allyl alcohol and malathion, described above.

Performance of DA when evaluated against MoA

Each substance was assigned a MoA using the MechoA scheme and then designated into broader categories of nonpolar narcotics, polar narcotics, reactive (specific), and reactive (unspecific). This categorization helped analyze the predictive performance of the DA, because several MechoA classifications occurred only once or twice and therefore could not be reliably assessed. The MoA were also assigned using the Verhaar classification scheme, but this resulted in an increased number of unclassified substances, which has been reported elsewhere (Firman et al., 2022). Therefore, the MechoA scheme was used for further evaluation. This comparison is reported in online supplementary material S5.

The predictions from the DA were evaluated per MoA, and very little difference was observed between the accuracy, sensitivity, specificity, and balanced accuracy (Figure 6). This analysis was undertaken on the subset of 49 substances with DA predictions derived from RTgill-W1 results and predictions from all three (Q)SARs, which led to relatively small numbers of substances to assess the impact of MoA. Therefore, it is difficult to make conclusive interpretations. However, it is interesting to note that



	All MoA (*n* =49)	Non-polar narcotics (*n* =20)	Polar narcotics (*n* =2)	Reactive (specific; *n* =11)	Reactive (unspecific; *n* =15)
Accuracy	79.6%	70.0%	100.0%	90.9%	80.0%
Sensitivity	81.3%	76.8%	100.0%	94.4%	83.8%
Specificity	93.1%	88.6%	100.0%	97.7%	94.6%
Balanced Accuracy	87.2%	82.7%	100.0%	97.2%	90.5%

Figure 6. The performance of the defined approach when evaluated against different broad categories of Mode of Action (MoA). Total $n = 48$ because one chemical could not be assigned by MechoA.

specific acting chemicals are predicted by the DA with over 90% accuracy because these are often difficult to predict with (Q)SARs. As additional RTgill-W1 data is generated, the DA can be further assessed against MoA and a more accurate representation will be possible.

Use of the geometric mean of the LC50 data for substances with multiple fish studies

When compiling and curating a large dataset, there are several mathematical ways to derive a "final" value when there are multiple studies with differing results. The median value may allow significant outliers (at either end) to have less influence. The mean provides a simple average value across all studies, whereas the geometric mean is often considered to be more accurate when there are diversity/outliers in a given dataset. In addition, when using biological data, it is common to use the most sensitive value (the lowest value), to be conservative in an evaluation of the substance. The GHS guidance states that typically the most sensitive value is used, and it is advised not to combine data that crosses different species in a taxonomic group, but that has to be considered on a case-by-case basis. Because there tends to be some flexibility in the guidance, and considering that for this project the resources were not available to check every single study to assess the reliability of the outlier values, it was therefore agreed to use the geometric mean. The geometric mean provides a better representation of all available data for a specific substance without undue influence from a minimum (conservative) value that could be an outlier. However, it is clear that species sensitivity for specific chemical classes needs to be a consideration for acute fish toxicity, as described for malathion previously. Despite this, in a real-world scenario, if an acute fish toxicity test was to be conducted on a new substance, it would be likely to only be tested once, and in a single species; thus, multiple values from discrete studies would not exist for comparison.

Choice of (Q)SARs and number of (Q)SARs used in the DA

The 12 (Q)SARs considered for use in the DA were those available from ECOSAR, VEGA, and T.E.S.T. These were chosen because they represented well-established, well-used, and easily available in silico models for the acute fish toxicity endpoint.

The evaluation of the 12 (Q)SARs, after removal of out of domain (Q)SAR predictions, illustrated that many had similar predictivity, which made choosing specific (Q)SARs to take forward for use in the DA challenging (Figure 3). A threshold of 75% for balanced accuracy was used initially, which left six potential (Q)SARs for use in the DA: USEPA ECOSAR, VEGA Fathead Minnow (USEPA), VEGA Fathead Minnow (KNN-IRFMN), VEGA Fish (KNN-Read-Across), VEGA Fish (NIC), VEGA Guppy (KNN-IRFMN), and T.E.S.T. (consensus).

Next, the performance metrics of the individual (Q)SARs per GHS class were evaluated. This illustrated that several of the six (Q)SARs with an overall high BA showed an unbalanced predictivity of each individual GHS class (e.g., Guppy [KNN-IRFMN] predicts Acute 1 much less well than it predicts Acute 2, Acute 3, or Not Classified substances—33% sensitivity versus >65%; Figure 4). Consequently, a by-class sensitivity threshold of 60% was applied to the 12 (Q)SARs under evaluation, and the three remaining (Q)SARs were ECOSAR, VEGA Fathead Minnow (KNN-IRFMN), and VEGA Fish (KNN-Read-Across), each with an overall balanced accuracy more than 75% and a by-class sensitivity more than 60%. These three (Q)SARs were selected for use in the DA for acute fish toxicity.

In total, there were four information sources used: predictions from USEPA ECOSAR, VEGA Fathead Minnow (KNN-IRFMN), and

VEGA Fish (KNN-Read-Across; Q)SARs and the output from the RTgill-W1 assay. When all of these results are applicable, the accuracy of the DA is 80%. However, when using either two or three (Q)SARs, in combination with the RTgill-W1 assay, the accuracy of the DA is similar at 78%. Consequently, to use the DA, reliable and in domain predictions must be available for at least two of the three recommended (Q)SARS, and a result from the RTgill-W1 assay must always be used. If fewer (Q)SAR predictions are available, or it is not possible to conduct the RTgill-W1 assay, then the scoring method used in the DA may still be used, but only as part of a wider weight of evidence approach.

The DA proposed through execution of this project is an example of a proof-of-concept that the VEGA Fathead Minnow (KNN-IRFMN), VEGA Fish (KNN-Read-Across), and USEPA ECOSAR (Q)SARs and RTgill-W1 data can be used in combination to predict GHS categories of acute fish toxicity. It is possible that similar accuracy could be obtained using different selections of (Q)SARs, especially if using another dataset to evaluate the performance of the set of 12 (Q)SARs.

Scoring used in the DA

When considering how to develop a DA for acute fish toxicity that could predict acute environmental GHS classification categories, the DA for skin sensitization known as the ITS was used as inspiration, because this DA also provides a categorical output suitable for use in EU CLP and/or GHS classification context (Organisation for Economic Co-Operation and Development, 2023; Takenouchi et al., 2015). This DA uses a score-based method to derive its predictions, so a similar method was undertaken for the DA for acute fish toxicity. Given that the goal of the project was to accurately predict GHS categories, converting each (Q)SAR prediction and RTgill-W1 outcome to their corresponding GHS classifications and assigning scores to each was used as a starting point. After several analyses testing out varying scoring methods, a simple approach of assigning scores to each information source, based on potency, calculating the mean score of the information sources, then converting back to a corresponding GHS category, provided accurate predictions. The score assigned to Acute 1 (score = 4) was designed to result in more conservative DA predictions by more heavily weighting an Acute 1 classification than Acute 2, Acute 3, or Not Classified classifications.

By-class performance of the DA

The DA demonstrates high accuracy in detecting both the most toxic and least toxic substances as illustrated by its accuracy for Acute 1 toxicants and Not Classified chemicals (90.9% and 87.5%, respectively). The DA is slightly less predictive for Acute 2 and Acute 3 substances (73.3%, for both) (Table 3). This reduced accuracy for moderately toxic substances is expected, because predicting moderate toxicity is generally more challenging.

Selected FP and FN case studies

Several false positive (substances predicted to be more toxic by the DA than indicated by the GHS category corresponding to the experimental fish data in the AFSA dataset) and FN (substances predicted to be Not Classified or less toxic by the DA than indicated by the GHS category corresponding to the experimental fish data in the AFSA dataset) were investigated further. For the presented case studies, it could be proposed that misclassifications by the DA can be attributed to potential in vivo outliers (dichloromethane) or borderline in vivo or (Q)SAR predictions/in vitro results (3-isobutyl-1-methylcyclohexan-1-ol (rossitol), cyclohexyl salicylate, di-n-butylorthophthalate). These are described in more detail in online supplementary material S4.

Case study 1

Chemical name: 3-isobutyl-1-methylcyclohexan-1-ol (rossitol)
DA prediction: Acute 3
Fish GHS category: Acute 2
FN/FP?: FN
Comment: Fish LC50 and two of the four DA information sources (Fathead Minnow [KNN-IRFMN] and Fish [KNN-Read-Across]) are clustered around the cutoff between Acute 2 and Acute (10 mg/L). These potentially borderline results increase uncertainty.

Case study 2

Chemical name: dichloromethane
DA prediction: Not Classified
Fish GHS category: Acute 3
FN/FP?: FN
Comment: Nine fish LC50s were reported, ranging from 2.6 to 502 mg/L. Seven of the nine LC50s are reported as >100 mg/L (Not Classified), but the presence of two LC50s at 2.6 and 2.9 mg/L resulted in the geometric mean being calculated as 91.5 mg/L, leading to an in vivo classification of Acute 3.

Case study 3

Chemical name: cyclohexyl salicylate
DA prediction: Acute 1
Fish GHS category: Acute 2
FN/FP?: FP
Comment: The heavier weighting of (Q)SAR predictions corresponding to Acute 1, combined with the rounding up of the mean DA score (mean = 2.5, rounded up to 3 = Acute 2 DA prediction), led to the misclassification of this substance. In addition, the fish LC50 was reported as 1.5 mg/L, close to the borderline between Acute 1 and Acute 2.

Case study 4

Chemical name: di-n-butylorthophthalate
DA prediction: Acute 1
Fish GHS category: Acute 2
FN/FP?: FP
Comment: Twenty fish LC50s were reported, ranging from 0.48 to 6.5 mg/L. Fifteen of the 20 LC50s are reported as >1 mg/L (Acute 2) and five as Acute 1. Three of the four information sources used in the DA assigned the substance as an Acute 1 toxin. These results all cluster relatively close to the border between Acute 1 and Acute 2.

Limitations of the acute fish toxicity DA

It is important to understand that this DA for acute fish toxicity is a proof-of-concept example of how non-animal information sources can be combined to provide reliable predictions of the GHS category of acute fish toxicity. The reason to highlight it as a proof-of-concept is because this DA has several limitations. The first is related to the quality of experimental data: The experimental fish LC50 data has been curated as described in the *Materials and Methods* section. It is likely that due to human error, there will be minor mistakes in the published source data, issues with study reliability, and other problems that have not been addressed by the curation and standardization undertaken in

this project, leading to variability in the experimental data used to benchmark the DA. This cannot be addressed without undertaking an in-depth, manual, labor-intensive substance by substance curation of the AFSA dataset. Species sensitivity is another limitation; as described previously for malathion, there are certain chemical classes and/or mechanisms of action that have varying sensitivities to different species of fish, which can impact the performance of the DA. Future iterations of the DA could provide species-specific or chemical class-specific modules. In addition, the DA has not been validated against a test set of unseen data; thus, it may only be applicable to substances that fall within the chemical space and MoA profiles observed in the AFSA dataset and described in this article. In addition, the DA is limited by the applicability domain of the (Q)SARs; newer chemistries may be less well represented in the training set of models and may be more likely to provide a lower reliability prediction, not suitable for use in the DA. Finally, a significant limitation of the DA is related to the accepted inputs into the (Q)SAR model. Many models cannot predict for mixtures, nor for metals or organometallics. Consequently, the DA described is only currently applicable for single organic chemicals. Mixtures could potentially be assessed using the DA (Q)SARs by inputting the individual components of a mixture and taking the most conservative prediction; however, this does not account for any synergistic effects. In addition, it may be possible to adapt the DA to use (Q)SARs that are able to provide reliable predictions for mixtures and/or unknown or variable composition, complex reaction products or of biological materials (UVCBs) or apply other calculations to account for intermolecular interactions (Bicherel & Thomas, 2021), because the RTgill-W1 assay is already applicable to mixtures and unknown or variable composition, complex reaction products or of biological materials (UVCBs).

Conclusions and further work

In conclusion, a proof-of-concept defined approach (DA) for acute fish toxicity has been developed that integrates data from four information sources: three freely available (Q)SARs, VEGA Fish (KNN-Read-Across), VEGA Fathead Minnow (KNN-IRFMN), and ECOSAR, and the in vitro RTgill-W1 assay. It provides GHS classifications for Acute 1, Acute 2, Acute 3, or Not Classified, making it suitable for use within EU CLP and/or GHS regulatory frameworks.

With respect to performance, this DA demonstrates an overall accuracy of 80% in predicting acute fish toxicity classifications as defined by GHS. The DA can also be used when only two (Q)SAR predictions are available. The DA is applicable to a wide range of chemical substances with diverse MoAs, as shown by analyses of chemical space coverage and MoA impact. However, there are several limitations associated with the DA that should be taken into account before use. For example, it is only suitable for single organic chemicals, and it has not been evaluated against an external test set.

A KNIME workflow and an example Excel file is available from the corresponding author. This workflow applies the scoring to each information source and provides the predicted GHS classification.

Further work to evaluate the DA against an unseen test set of data could be undertaken and could lead to amendment of the score-based approach to improve its applicability and scope of use.

Supplementary material

Supplementary material is available online at *Environmental Toxicology and Chemistry*.

Data availability

The Animal-Free Safety Assessment dataset can be accessed in the online supplementary material.

Author contributions

Donna S. Macmillan (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Visualization, Writing—original draft, Writing—review & editing), Pravin Ambure (Conceptualization, Data curation, Formal analysis), Vicente Aranda (Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing—original draft, Writing—review & editing), Yannick Bayona (Conceptualization, Writing—original draft, Writing—review & editing), Vera Bonderovic (Conceptualization, Writing—review & editing), Jay Dawick (Conceptualization, Writing—review & editing), Nicolas Fabre (Conceptualization, Writing—original draft, Writing—review & editing), Stephan Fischer (Conceptualization, Writing—review & editing), Geoff Hodges (Conceptualization, Investigation, Writing—original draft, Writing—review & editing), Ágata Llobet-Mut (Conceptualization, Data curation, Formal analysis, Methodology), Sophie Loisel-Joubert (Conceptualization, Writing—review & editing), Claudia Rivetti (Conceptualization, Writing—review & editing), Jayne Roberts (Conceptualization, Investigation, Writing—original draft, Writing—review & editing), Kristin Schirmer (Conceptualization, Writing—review & editing), Eva Serrano Candelas (Conceptualization, Data curation, Formal analysis, Methodology, Resources, Software, Visualization, Writing—original draft, Writing—review & editing), Blanca Serrano Ramón (Conceptualization, Writing—review & editing), and Ricky A. Stackhouse (Conceptualization, Writing—review & editing)

Funding

None declared.

Conflicts of interest

None declared.

Ethics statement

No animals used.

References

Alépée, N., & Adriaens, E. (2024). Development of a defined approach for eye hazard identification of solid chemicals according to the three UN GHS categories. *ALTEX*, 41, 457–468. https://doi.org/10.14573/altex.2401191

Alépée, N., Adriaens, E., Abo, T., Bagley, D., Desprez, B., Hibatallah, J., Mewes, K., Pfannenbecker, U., Sala, À., Van Rompay, A. R., Verstraelen, S., & McNamee, P. (2019). Development of a defined approach for eye irritation or serious eye damage for neat liquids based on cosmetics Europe analysis of *in vitro* RhCE and BCOP test methods. *Toxicology in Vitro*, 59, 100–114. https://doi.org/10.1016/j.tiv.2019.04.011

Alépée, N., Adriaens, E., Abo, T., Magby, J., Mewes, K. R., & Giusti, A. (2023). Development of a Defined Approach for Eye hazard identification of chemicals having surfactant properties according to the three UN GHS categories. *Toxicology in Vitro*, 89, 105576. https://doi.org/10.1016/j.tiv.2023.105576

Animal-Free Safety Assessment Collaboration. (2018). *Animal-Free Safety Assessment (AFSA) Collaboration*. https://www.afsacollaboration.org/

Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. *ACM SIGMOD Record*, 28, 49–60. https://doi.org/10.1145/304181.304187

Bauer, F., Thomas, P., Fouchard, S., & Serge, N. (2018). A new classification algorithm based on Mechanisms of Action. *Computational Toxicology*, 5, 8–15. https://doi.org/10.1016/j.comtox.2017.11.001

Benfenati, E., Manganaro, A., & Gini, G. (2013). VEGA-QSAR: AI inside a platform for predictive toxicology. *CEUR Workshop Proceedings*, 1107, 21–28.

Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Thiel, K., & Wiswedel, B. (2009). KNIME—the Konstanz information miner: Version 2.0 and beyond. *ACM SIGKDD Explorations Newsletter*, 11, 26–31. https://doi.org/10.1145/1656274.1656280

Bicherel, P., & Thomas, P. C. (2021). Aquatic toxicity calculation of mixtures: A chemical activity approach incorporating a bioavailability reduction concept. *Environmental Science & Technology*, 55, 11183–11191. https://doi.org/10.1021/acs.est.1c04389

Cronin, M. T. D. (2017). (Q)SARs to predict environmental toxicities: Current status and future needs. *Environmental Science. Processes & Impacts*, 19, 213–220. https://doi.org/10.1039/C6EM00687F

Danieli, A., Colombo, E., Raitano, G., Lombardo, A., Roncaglioni, A., Manganaro, A., Sommovigo, A., Carnesecchi, E., Dorne, J.-L. C. M., & Benfenati, E. (2023). The VEGA tool to check the applicability domain gives greater confidence in the prediction of in silico models. *International Journal of Molecular Sciences*, 24 Article 12. https://doi.org/10.3390/ijms24129894

Drugs.com. (2024). *List of nutraceutical products*. Drugs.Com. https://www.drugs.com/drug-class/nutraceutical-products.html

European Chemicals Agency. (2008). *Guidance on information requirements and chemical safety assessment—Chapter R.6: QSARs and grouping of chemicals*. https://echa.europa.eu/documents/10162/17224/information_requirements_r6_en.pdf/77f49f81-b76d-40ab-8513-4f3a533b6ac9

European Chemicals Agency. (2016). *Practical guide—How to use and report (Q)SARs*. Publications Office. https://data.europa.eu/doi/10.2823/81818

European Chemicals Agency. (2023). *The use of alternatives to testing on animals for the REACH Regulation*. Publications Office. https://data.europa.eu/doi/10.2823/805454

European Commission. (2022). Alures—animal use reporting—EU statistics database on the use of animals for scientific purposes under directive 2010/63/EU. https://webgate.ec.europa.eu/envdataportal/content/alures/section2_number-of-uses.html

European Union. (2006). Regulation (EC) No. 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No. 793/93 and Commission Regulation (EC) No. 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. *Official Journal of the European Union*, 49, 1–396. https://eur-lex.europa.eu/eli/reg/2006/1907/oj

European Union. (2008). Regulation (EC) No. 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging. *Official Journal of the European Union*, 51, 1–353. https://eur-lex.europa.eu/eli/reg/2008/1272/oj

European Union. (2009). Regulation (EC) No. 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. *Official Journal of the European Union*, 52, 1–309. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:309:0001:0050:en:PDF

European Union. (2010). Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes. *Official Journal of the European Union*, 53, 1–276. https://eur-lex.europa.eu/eli/dir/2010/63/oj/eng

European Union. (2019). Regulation (EU) 2019/1010 of the European Parliament and of the Council of 5 June 2019 on the alignment of reporting obligations in the field of legislation related to the environment, and amending Regulations (EC) No. 166/2006 and (EU) No. 995/2010 of the European Parliament and of the Council, Directives 2002/49/EC, 2004/35/EC, 2007/2/EC, 2009/147/EC and 2010/63/EU of the European Parliament and of the Council, Council Regulations (EC) No. 338/97 and (EC) No. 2173/2005, and Council Directive 86/278/EEC (Text with EEA relevance). http://data.europa.eu/eli/reg/2019/1010/oj/eng

Firman, J. W., Ebbrell, D. J., Bauer, F. J., Sapounidou, M., Hodges, G., Campos, B., Roberts, J., Gutsell, S., Thomas, P. C., Bonnell, M., & Cronin, M. T. D. (2022). Construction of an in silico structural profiling tool facilitating mechanistically grounded classification of aquatic toxicants. *Environmental Science & Technology*, 56, 17805–17814. https://doi.org/10.1021/acs.est.2c03736

Fischer, M., Belanger, S. E., Berckmans, P., Bernhard, M. J., Bláha, L., Coman Schmid, D. E., Dyer, S. D., Haupt, T., Hermens, J. L. M., Hultman, M. T., Laue, H., Lillicrap, A., Mlnaříková, M., Natsch, A., Novák, J., Sinnige, T. L., Tollefsen, K. E., von Niederhäusern, V., Witters, H., , … Schirmer, K. (2019). Repeatability and reproducibility of the RTgill-W1 cell line assay for predicting fish acute toxicity. *Toxicological Sciences: An Official Journal of the Society of Toxicology*, 169, 353–364. https://doi.org/10.1093/toxsci/kfz057

FooDB. (2024). *FooDB*. https://foodb.ca/

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2, 1–21. https://doi.org/10.1038/s43586-022-00184-w

International Fragrance Association. (2024). *IFRA Transparency List*. https://ifrafragrance.org/priorities/ingredients/ifra-transparency-list

International Union of Pure and Applied Chemistry. (1997a). *MOL file format*. https://doi.org/10.1351/goldbook.MT06966

International Union of Pure and Applied Chemistry. (1997b). *SMILES*. https://doi.org/10.1351/goldbook.ST06980

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2023). PubChem 2023 update. *Nucleic Acids Research*, 51, D1373–D1380. https://doi.org/10.1093/nar/gkac956

KREATiS (2024). *iSafeRat*. https://www.kreatis.eu/

Lee, Y.-L., Chen, Z.-Y., Li, T.-N., Hsueh, J.-F., & Wang, Y.-J. (2024). A novel integrated testing strategy (ITS) for evaluating acute fish toxicity with new approach methodologies (NAMs). *Environment*

International, 193, 109112. https://doi.org/10.1016/j.envint.2024.109112

Lillicrap, A., Moe, S. J., Wolf, R., Connors, K. A., Rawlings, J. M., Landis, W. G., Madsen, A., & Belanger, S. E. (2020). Evaluation of a Bayesian network for strengthening the weight of evidence to predict acute fish toxicity from fish embryo toxicity data. *Integrated Environmental Assessment and Management*, 16, 452–460. https://doi.org/10.1002/ieam.4258

Martin, T., Harten, P., & Young, D. (2023). *TEST (Toxicity Estimation Software Tool)* (Version 5.1.2.0.) [Computer software]. https://www.epa.gov/comptox-tools/toxicity-estimation-software-tool-test

Morgan, H. L. (1965). The generation of a unique machine description for chemical structures-a technique developed at chemical abstracts service. *Journal of Chemical Documentation*, 5, 107–113. https://doi.org/10.1021/c160017a018

National Institutes of Health. (2024a). *NCI/CADD Chemical Identifier Resolver*. https://cactus.nci.nih.gov/chemical/structure

National Institutes of Health. (2024b). *PubChem*. https://pubchem.ncbi.nlm.nih.gov/

Natsch, A., Laue, H., Haupt, T., von Niederhäusern, V., & Sanders, G. (2018). Accurate prediction of acute fish toxicity of fragrance chemicals with the RTgill-W1 cell assay. *Environmental Toxicology and Chemistry*, 37, 931–941. https://doi.org/10.1002/etc4027

Organisation for Economic Co-Operation and Development. (2004). *Test No. 202: Daphnia sp. Acute Immobilisation Test*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/environment/test-no-202-daphnia-sp-acute-immobilisation-test_9789264069947-en

Organisation for Economic Co-Operation and Development. (2006). *Test No. 221: Lemna sp. Growth Inhibition Test*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/environment/test-no-221-lemna-sp-growth-inhabition-test_9789264016194-en

Organisation for Economic Co-Operation and Development. (2011). *Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/environment/test-no-201-alga-growth-inhibition-test_9789264069923-en

Organisation for Economic Co-Operation and Development. (2013). *Test No. 236: Fish Embryo Acute Toxicity (FET) Test*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/environment/test-no-236-fish-embryo-acute-toxicity-fet-test_9789264203709-en

Organisation for Economic Co-operation and Development. (2014). *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*. OECD Publishing. https://doi.org/10.1787/9789264085442-en.

Organisation for Economic Co-operation and Development. (2017a). *Guidance Document for the Use of Adverse Outcome Pathways in Developing Integrated Approaches to Testing and Assessment (IATA)*. OECD Publishing.

Organisation for Economic Co-operation and Development. (2017b). *Guidance Document on the Reporting of Defined Approaches and Individual Information Sources to be Used within Integrated Approaches to Testing and Assessment (IATA) for Skin Sensitisation*. OECD Publishing.

Organisation for Economic Co-operation and Development. (2019). *Second Edition—Guidance Document on Integrated Approaches to Testing and Assessment (IATA) for Serious Eye Damage and Eye Irritation*. OECD Publishing.

Organisation for Economic Co-Operation and Development. (2019). *Test No. 203: Fish, Acute Toxicity Test*. OECD Publishing. https://www.oecd-ilibrary.org/environment/test-no-203-fish-acute-toxicity-test_9789264069961-en

Organisation for Economic Co-Operation and Development. (2021a). *OECD QSAR Toolbox* (Version 4.5) [Computer software]. https://qsartoolbox.org/

Organisation for Economic Co-Operation and Development. (2021b). *Test No. 249: Fish Cell Line Acute Toxicity—The RTgill-W1 cell line assay*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/environment/test-no-249-fish-cell-line-acute-toxicity-the-rtgill-w1-cell-line-assay_c66d5190-en

Organisation for Economic Co-Operation and Development. (2023). *Guideline No. 497: Defined Approaches on Skin Sensitisation*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/environment/guideline-no-497-defined-approaches-on-skin-sensitisation_b92879a4-en

Organisation for Economic Co-operation and Development. (2024). *Guideline No. 467: Defined Approaches for Serious Eye Damage and Eye Irritation*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/environment/test-no-467-defined-approaches-for-serious-eye-damage-and-eye-irritation_28fe2841-en

ProtoQSAR. (2025). *ProtoPRED* [Computer software]. https://protopred.protoqsar.com/

Python Software Foundation. (2024). *Python Language Reference, version 3.9.4. Available at* http://www.python.org. (Version 3.9.4) [Computer software]. http://www.python.org

RDKit: Open-Source Cheminformatics. (2021). *RDKit: Open-source cheminformatics*. https://www.rdkit.org (Version 2021.03.2) [Computer software]. https://www.rdkit.org

Reed, N. R., & Rubin, A. L. (2014). Malathion. In P. Wexler (Ed.), *Encyclopedia of Toxicology* (3rd ed., pp. 133–137). Academic Press. https://doi.org/10.1016/B978-0-12-386454-3.00156-1

Reuschenbach, P., Silvani, M., Dammann, M., Warnecke, D., & Knacker, T. (2008). ECOSAR model performance with a large test set of industrial chemicals. *Chemosphere*, *71*, 1986–1995. https://doi.org/10.1016/j.chemosphere.2007.12.006

Russell, W. M. S., & Burch, R. L. (1960). The principles of humane experimental technique. *Medical Journal of Australia*, 1, 500. https://doi.org/10.5694/j.1326-5377.1960.tb73127.x

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., & Steinbeck, C. (2021). COCONUT online: Collection of Open Natural Products database. *Journal of Cheminformatics*, 13, 2. https://doi.org/10.1186/s13321-020-00478-9

Sterling, T., & Irwin, J. J. (2015). ZINC 15—Ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55, 2324–2337. https://doi.org/10.1021/acs.jcim.5b00559

Strickland, J., Truax, J., Corvaro, M., Settivari, R., Henriquez, J., McFadden, J., Gulledge, T., Johnson, V., Gehen, S., Germolec, D., Allen, D. G., & Kleinstreuer, N. (2022). Application of defined approaches for skin sensitization to agrochemical products. *Frontiers in Toxicology*, 4, 852856. https://doi.org/10.3389/ftox.2022.852856

Takenouchi, O., Fukui, S., Okamoto, K., Kurotani, S., Imai, N., Fujishiro, M., Kyotani, D., Kato, Y., Kasahara, T., Fujita, M., Toyoda, A., Sekiya, D., Watanabe, S., Seto, H., Hirota, M., Ashikaga, T., & Miyazawa, M. (2015). Test battery with the human cell line activation test, direct peptide reactivity assay and DEREK based on a 139 chemical data set for predicting skin

sensitizing potential and potency of chemicals. *Journal of Applied Toxicology: JAT*, 35, 1318–1332. https://doi.org/10.1002/jat.3127

Tanneberger, K., Knöbel, M., Busser, F. J. M., Sinnige, T. L., Hermens, J. L. M., & Schirmer, K. (2013). Predicting fish acute toxicity using a fish gill cell line-based toxicity assay. *Environmental Science & Technology*, 47, 1110–1119. https://doi.org/10.1021/es303505z

United Nations Globally Harmonized System. (2023). *Globally Harmonized System of Classification and Labelling of Chemicals (GHS Rev. 10, 2023) | UNECE*.https://unece.org/transport/dangerous-goods/ghs-rev10-2023

United States Environmental Protection Agency. (2013, February 22). *Summary of the Toxic Substances Control Act* [Overviews and Factsheets]. https://www.epa.gov/laws-regulations/summary-toxic-substances-control-act

United States Environmental Protection Agency. (2024). *ECOSAR* (Version 2.2) [Computer software]. https://www.epa.gov/tsca-screening-tools/ecological-structure-activity-relationships-ecosar-predictive-model

Voigt, M., & Jaeger, M. (2023). Chapter 36—In silico and in vivo ecotoxicity—QSAR-based predictions and experimental assays for the aquatic environment. In H. Hong (Ed.), *QSAR in Safety Evaluation and Risk Assessment* (pp. 495–509). Academic Press. https://doi.org/10.1016/B978-0-443-15339-6.00018-7

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., … Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46, D1074–D1082. https://doi.org/10.1093/nar/gkx1037

Zeng, X., Zhang, P., He, W., Qin, C., Chen, S., Tao, L., Wang, Y., Tan, Y., Gao, D., Wang, B., Chen, Z., Chen, W., Jiang, Y. Y., & Chen, Y. Z. (2018). NPASS: Natural product activity and species source database for natural product research, discovery and tool development. *Nucleic Acids Research*, 46, D1217–D1222. https://doi.org/10.1093/nar/gkx1026

ZINC 15. (2024). *ZINC 15*. https://zinc15.docking.org/catalogs/dbfda/substances/subsets/world/

Environmental Toxicology and Chemistry, 2025, 00, 1–14
https://doi.org/10.1093/etojnl/vgaf046
Original Article