Evaluation of the OECD QSAR toolbox automatic workflow for the prediction of the acute toxicity of organic chemicals to fathead minnow

Enrico Mombelli, Pascal Pandard

Evaluation of the OECD QSAR Toolbox automatic workflow for the prediction of the acute toxicity of organic chemicals to fathead minnow

Enrico Mombelli[a, *], Pascal Pandard[a]

[a]Institut National de l'Environnement Industriel et des Risques (INERIS), 60550 Verneuil en Halatte, France

* Corresponding author

E-mail address: enrico.mombelli@ineris.fr (E. Mombelli)

Abstract

Regulatory frameworks require information on acute fish toxicity to ensure environmental protection. The experimental assessment of this property relies on a substantial number of fish to be tested and it is in conflict with the current drive to replace *in vivo* testing. For this reason, alternatives to *in vivo* testing have been proposed during the past years. Among these alternatives, there are Quantitative Structure-Activity Relationships (QSAR) that require the sole knowledge of chemical structure to yield predictions of toxicities. In this context, the OECD QSAR Toolbox is one of the leading QSAR tools for regulatory purposes that enables the prediction of fish toxicities. The aim of this work is to provide evidence about the predictive reliability of the automated workflow for predicting acute toxicity in fish embedded within this toolbox. The results herein presented show that the logic underpinning this automated workflow can predict with a reliability that, in the majority of cases, is comparable to inter-laboratory variability and, in a significant number of cases, is also comparable with intra-laboratory variability. Moreover, considerations on the toxic mode of action provided by the OECD tool proved to be helpful in refining predictions and reducing the number of prediction outliers.

Keywords: QSAR, Fish toxicity, 3Rs, OECD QSAR Toolbox, automated workflow

1. Introduction

Regulations on chemical toxicity aim at protecting the environment from the adverse effects caused by exposure to chemicals which can threaten the trophic chain in aquatic ecosystems (McCarty et al., 2018). In this respect, the protection of the trophic chain that, from algae leads to fish through intermediate organism (e.g., crustaceans), is of great importance for the preservation of ecosystems and, indirectly, it can also protect humans from dangerous intoxications (Barletta and Lima, 2019). Therefore, an understanding of toxic effects at representative trophic levels is of great importance in regulatory settings and fish represent one of the closest ecological sentinels to humans whose fish consumption is characterized by an ever-increasing growth-rate (FAO, 2020).

The regulatory importance of testing in fish can be illustrated by the European regulation REACH that requires acute fish testing for chemicals produced or imported in quantities larger than 10 tons per year (EC, 2006). Fish acute toxicity is generally accomplished by following the recommendations reported in the OECD guideline 203 (OECD, 2019). According to this guideline, an *in vivo* test is carried out by exposing fish for 96 hours to estimate the concentration resulting in 50% of fish lethality (LC_{50}). This experimental protocol requires, at least, 42 fish to be used in a single experiment if a full test has to be performed (Schug et al., 2020).

A threshold approach has been developed to reduce the number of animals (OECD, 2010). This testing strategy recommends performing a limit test for fish acute toxicity at a threshold concentration (TC) corresponding to the lowest EC_{50} obtained in the algal growth test or the daphnia acute toxicity test. If no mortality occurs in the limit test, fish are less sensitive than the species from the other taxonomic groups and the TC is used as a surrogate of the LC_{50} value in hazard/risk assessment or for classification and labelling purposes. Computational approaches can also represent reliable alternatives to *in vivo* testing in fish (Netzeva et al., 2008). These computational methods, also known as *in silico* approaches, rely on Quantitative Structure-Activity

Relationships (QSAR) that can rapidly predict biological properties (e.g., lethal concentrations) as a function of chemical structure (Cherkasov et al., 2014). The key-tenet of QSAR approaches is that similar chemicals induces similar (qualitatively and quantitively) effects in living beings (Cherkasov et al., 2014). One of the most simple and effective paradigms of a QSAR approach in ecotoxicity is represented by the seminal work of Könemann (Könemann, 1981) that successfully related a change in toxicological potency of chemicals in fish to a change in the octanol-water partition coefficient (Log P, a molecular property or descriptor) that characterizes chemical substances. In other words, Könemman showed that chemicals with similar Log P values are characterized by similar LC_{50} values.

The QSAR paradigm is endorsed by one of the most widely used *in silico* tools in regulatory settings: the OECD QSAR toolbox (Dimitrov et al., 2016; OECD, 2020). The OECD QSAR Toolbox (hereafter referred to as "the Toolbox") is a freely available *in silico* system which contains several databases to provide information on toxicological, ecotoxicological and physicochemical properties. The Toolbox enables users to group chemicals within categories based on mechanistic (i.e. similar mode of action) and structural similarities (e.g., same organic functional groups) and then predict toxicological properties thanks to the QSAR paradigm that translates into trend analysis and read-across predictions (Dimitrov et al., 2016).

The aim of the work described in this paper is to evaluate the predictive performance of the automated workflow (AW) for the prediction of LC_{50} at 96 hours for fathead minnow (*Pimephales promelas*). An AW is an algorithm that computes a QSAR prediction with a minimal interaction from the user. In the specific case of the AW for the prediction of acute toxicity in fathead minnow, the only action which is needed from a user is the identification (e.g., by CAS RN or by name) of the chemical for which an experimental LC_{50} is not available (hereafter referred to as "target chemical"). After the definition of the target chemical, the AW oversees the identification of appropriate chemicals that can be considered structurally and mechanistically similar to the target

chemical (hereafter referred to as “structural analogs”) and that are associated with experimental LC_{50} values that will permit QSAR predictions.

Knowledge about the reliability of this AW, is an important piece of information within regulatory contexts, since it provides an insight into what can be expected in terms of predictivity when replacing *in vivo* testing in fish with automated QSAR predictions yielded by the Toolbox. By providing information on this operational aspect, this article will increase the awareness of end-users about what can be reasonably expected in terms of precision from the AW for the prediction of LC_{50} in fathead minnow.

2. Materials and methods

2.1 *The AW for the prediction of LC_{50} in fathead minnow*

The AW used in this article is those implemented in the Toolbox v 4.3 (OECD, 2020). This AW is thoroughly described in Yordanova et al. (Yordanova et al., 2019) and it can be briefly summarized as follows:

1) Input: the target chemical is defined by the user (e.g., by inputting its CAS RN).

2) Profiling: identification of structural features and mode of action (MOA) that characterize the target chemical applies a series of computerized modules (known as profilers) whose role is to identify structurally and mechanistic feature that are relevant for aquatic toxicity. These profiling results are then used as query criteria for recognizing appropriate structural analogues.

3) Data gathering: pertinent ecotoxicological databases are searched to retrieve a starting group of structural analogs that forms a broad initial category to be subsequently refined in terms of structural and mechanistic pertinence.

4) Subcategorization: information associated with non-discrete chemical entities (e.g., chemical mixtures) and LC_{50} greater than water solubility is discarded. After these first steps, the AW iteratively applies profilers to further enhance the structural and mechanistic coherence of the initial category of structural analogs.

5) Prediction: according to the number of identified structural analogs (NuA), this category of chemicals is then used to compute a LC_{50} prediction. If the number of structural analogs is greater than (or equal to) ten, a linear QSAR model that quantifies variations in LC_{50} values as a function of the octanol water partition coefficient Log K_{ow} is defined and used to compute a prediction (i.e. prediction by "trend analysis). Otherwise, a prediction is computed by estimating the arithmetic mean of the LC_{50} values that characterize the structural analogs. This second form of prediction is commonly referred to as prediction by read-across (van Leeuwen et al., 2009).

Before being communicated to the user, predictions are checked against criteria for acceptance (Yordanova et al., 2019):

- Trend analysis: if the coefficient of determination of the linear regression R^2 is ≥ 0.7 and NuA ≥ 10 THEN accept trend prediction ELSE switch to read-across
- Read-across: if the prediction corresponds to an interpolation and LC_{50} ≤ 2 log units OR log K_{ow} ≤ 2 log units AND NuA ≥ 5 THEN accept prediction and proceed with Report.

2.2 Experimental data on acute toxicity to fathead minnow

Experimental data on acute toxicity to fathead minnow were retrieved from the PubChem website and they correspond to the EPAFHM database (EPA, 2020). This database reports LC_{50} values for 617 compounds that were determined after 96 hours flow-through exposures using 28 to 36 days old

juveniles fathead minnows (Russom et al., 1997). The chemicals that compose this database describe a cross-section of industrial organic chemicals (Russom et al., 1997).

Exclusion of organometallic chemicals, inorganic chemicals, organic salts and mixtures gave rise to a dataset of 96 hours LC_{50} data for 553 discrete organic chemicals.

2.3 Partition Around Medoids

Given the large number of chemicals composing this database and the time required to obtain a prediction from the AW, the evaluation was carried out on a selected subset of chemicals. A structurally representative subset of chemicals was extracted from the EPAFHM database by using the Partition Around Medoids (PAM) algorithm (Wehrens, 2011).

A medoid is identified as an object (chemicals in the case of this article) that occupies the center of a cluster of similar items, whose average dissimilarity to all the objects in the cluster is minimal (Wehrens, 2011). The PAM algorithm was applied to chemical descriptors computed thanks to the freely available software PaDEL v 2.21 (Yap, 2011) and the function pam of the *cluster* package v 2.1.0 (Maechler et al., 2019) for the free software environment R v3.6.1 (R Core Team, 2019). The optimal number of medoids k was identified by retaining the number of medoids that maximize the average of all silhouette widths. The highest average silhouette width indicates the optimal achievable clustering over a range of possible values for k. In the case of the present work we analyzed k values ranging from 2 to 150 and identified an optimal number of medoids equal to 145.

Medoids were visualized thanks to a Principal Component Analysis (PCA) executed by using the function prcomp of the R package stats v3.6.1 on centered and scaled descriptors.

2.4 *Evaluation strategy*

In the framework of the analysis described in this article the AW was prompted by inputting the CAS RN of chemicals. Moreover, one extra requirement on the selection of structural analogs was imposed: data with qualifiers (e.g. LC_{50} greater than a certain threshold) were removed each time that the AW flagged their presence.

Finally, it is important to highlight the fact that experimental data associated with the target chemical and present within the databases of the Toolbox were not considered when computing predictions by means of the AW. In other words, predictions were estimated as if experimental toxicities of target chemicals were unknown.

3. Results and Discussion

3.1 *Identification of an evaluation set of structurally representative chemicals*

The PAM algorithm described in the methodological section identified a subset of chemicals composed by 145 substances that can be regarded as structurally representative of the entire database. Indeed, as shown on Fig. 1, the medoids (black triangles) span in the plane defined by the first two principal components of a PCA analysis. This dataset was adopted as an evaluation set and it is composed by 145 chemicals whose LC_{50} values ranges from 0.0051 mg/L to 68900 mg/L. This evaluation set is available as Supplemental Information (Supplemental Table S1).

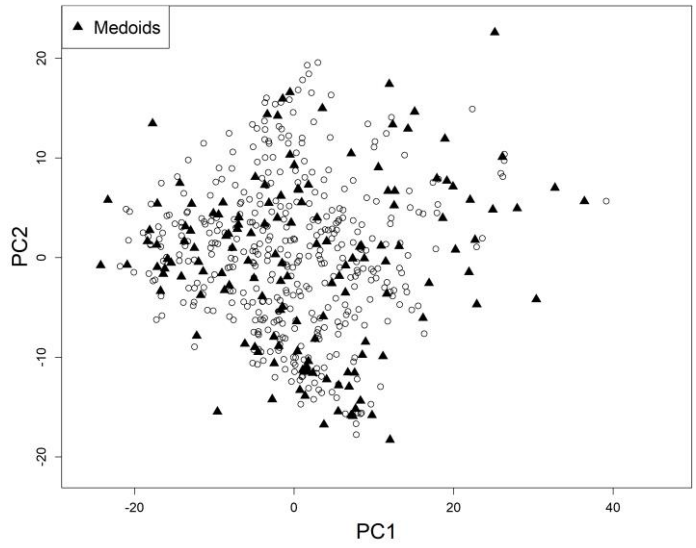

Fig. 1. Principal Component Analysis of the evaluation database. The medoids (black triangles) span the chemical space covered by the database.

3.2 *Experimental variability of 96h LC_{50} values*

When evaluating QSAR models it is generally assumed that their level of precision has to be benchmarked against the level of experimental reproducibility that characterizes the endpoint that the model predicts for (Cappelli et al., 2015; Cassano et al., 2014; Mombelli, 2012). According to an analysis made by Hrovat et al. (Hrovat et al., 2009) the difference between available values for the minimum and maximum 96 h LC_{50} associated with the same chemical can in several case be larger than one logarithmic unit. In the presence of such a large variability, all the QSAR predictions whose error is below a ten-fold factor could be regarded as reliable.

Therefore, to have a more precise idea about variability thresholds against which to benchmark QSAR predictions we searched ecotoxicological literature to find evidence about the variability that characterizes acute toxicity studies with fathead minnow.

The results of this bibliographic search, focusing on flow-through exposure, for consistency with the EPAFHM database are summarized in Table 1. Modes of toxicological action for organic chemicals and

inorganic chemicals (e.g. metals) differ but metals and other inorganic chemicals are routinely used, as reference chemicals to assess the sensitivity of the test organisms whatever the type of chemical to be tested (Römbke and Ahtiainen, 2007). For this reason, we deemed appropriate to include also data on metals when estimating experimental variability.

Studies on the variability of LC_{50} values generally report means and coefficients of variations of the underlying experimental data. Nevertheless, distributions of LC_{50} values are better described by log-normal distributions (e.g., LC_{50} values cannot be negative). For this reason, equations 1 and 2 that define the mean (m) and variance (v) of the variable's natural logarithm, were applied as a function of the mean M and variance V of the data reported in the publications cited in Table 1 (Johnson et al., 1994):

$$m = \ln\left(\frac{M^2}{\sqrt{M^2 + V^2}}\right) (Eq. 1)$$

$$v = \ln\left(1 + \frac{V}{M^2}\right) (Eq. 2)$$

These equations were useful for the estimation of a variability ratio that we defined as the ratio between the 97.5th and 2.5th percentiles of the logarithmic distribution.

Table. 1. Inter- and intra-laboratory reproducibility of 96h LC_{50} determined in fathead minnow by means of a flow-through exposure system. The variability ratio was estimated by computing the ratio between the 97.5th and 2.5th percentiles that characterizes the logarithmic distribution of 96h LC_{50}.

Chemical	Average LC_{50} [mg/L]	CV%	Variability ratio	Type of variability	Reference
Silver (as silver nitrate)	$7.49\ 10^{-3}$	40	4.5	Inter-laboratory	(EPA, 2002)
Endosulfan	$0.96\ 10^{-3}$	46	5.6	Inter-laboratory	(EPA, 2002)
Phenol	26.5	9.0	1.4	Inter-laboratory	(Walker, 1988)
Pentachlorophenol	0.21	12	1.6	Intra-laboratory	(Adelman et al., 1976)
Cadmium (as cadmium sulfate)	7.18	59	8.5	Intra-laboratory	(Pickering and Gast, 1972)
Hexavalent Chromium (as potassium dichromate)	48	22	2.3	Intra-laboratory	(Adelman et al., 1976)
Copper (as copper sulfate)	108	16	1.9	Intra-laboratory	(Lind et al., 1978)
Nickel (as nickel sulfate)	12.9	35.3	3.8	Intra-laboratory	(Lind et al., 1978)

An analysis of the references reported in Table 1 highlighted two investigations that we judged as being less reliable: the tests for phenol (Walker, 1988) and the tests for cadmium (Pickering and Gast, 1972). Indeed, the first study is characterized by a very narrow variability regardless of the heterogenous experimental conditions and also Environment-Canada highlighted this atypical issue (Environment-Canada, 1990). As far as the tests for cadmium are concerned (Pickering and Gast, 1972), it must be noted that the reported important precipitations of the test substance and pH variations in the water could have impaired the pertinence of the generated data.

For these reasons, we excluded data on phenol and cadmium when computing the geometric means of individual variability factors (Tab. 1) that defined the final intra and inter-laboratory variability ratios to be used as benchmarking references for the evaluation of the precision of QSAR predictions: 2.3 (intra-laboratory fold-factor) and 5.0 (inter-laboratory fold-factor).

3.3 *Evaluation of the AW*

The AW generated a prediction satisfying its acceptability criteria for 122 chemicals (90 trend predictions and 32 read-across predictions) out of 145 (supplemental Table S1). The plot depicting the observed 96 h experimental values vs. the corresponding predicted values is reported in Fig. 2. The prediction error associated with nine predictions (squares in Fig. 2) lies beyond a ten-fold factor (these LC_{50} are largely overpredicted), 85.3% of the predictions are characterized by an error which is below the inter-laboratory variability factor and 59.0% of the predictions are characterized by an error which is below the intra-laboratory variability factor.

The Concordance Correlation Coefficient (CCC) (Chirico and Gramatica, 2011) that characterizes all the prediction reported in Fig. 2 is equal to 0.90. This coefficient assesses precision and accuracy of predictions and any deviation from the bisector line representing perfect predictions results in a value of CCC which is smaller than 1 (Chirico and Gramatica, 2011). The predictivity of QSAR models is generally benchmarked against this statistical indicator and QSAR models are usually regarded as valid if the CCC is equal to or greater than 0.85.

Similarly, the coefficient of determination R^2 that characterizes these predictions is equal to 0.83 and it is usually recommended to ascertain if this statistical indicator exceeds a value of 0.7 to consider QSAR models as valid (Chirico and Gramatica, 2011). Therefore, it appears that, according to generally adopted quality standards, the AW can be regarded as valid and characterized by a performance which is comparable to inter-laboratory variability for the majority of cases.

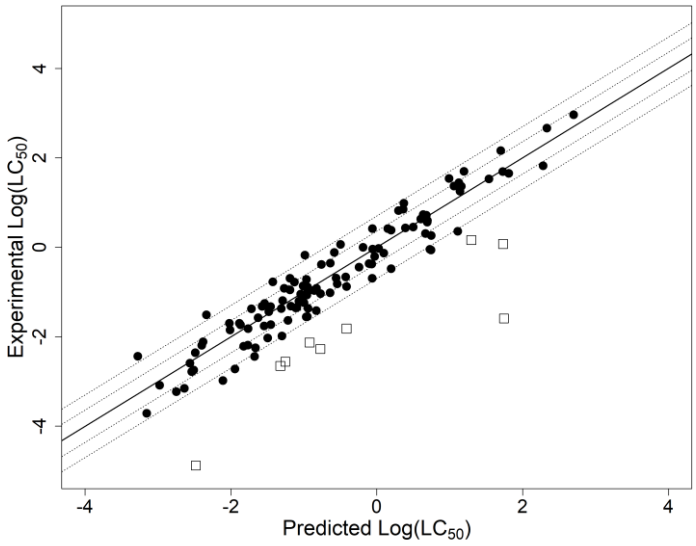

Fig. 2. Experimental vs. predicted 96 h LC_{50} values for fathead minnow. The dotted lines represent the intra- and inter-laboratory variability factors (2.3 and 5 respectively). Squares indicate extreme prediction outliers (error > 10-fold factor). The logarithmic values refer to concentrations expressed in mM on a linear scale.

The evaluation was repeated by retaining only the structural analogs associated with data from the EPAFHM database (EPA, 2020) to have a better insight into a benchmarking with respect to intra-laboratory variability. As described above, experimental data for the target chemicals were ignored when applying the AW. This second evaluation (Supplemental Table S2) was characterized by 96 predictions satisfying the acceptability of the AW: 70 trend predictions and 26 read-across predictions.

The plot depicting the observed 96 h experimental values vs. the corresponding predicted values for this second evaluation is reported in Fig. 3. The prediction error associated with seven predictions (squares in Fig. 3) lies beyond a ten-fold factor (six LC_{50} are largely overpredicted), 56.3% of the predictions are characterized by an error which is below the intra-laboratory variability factor.

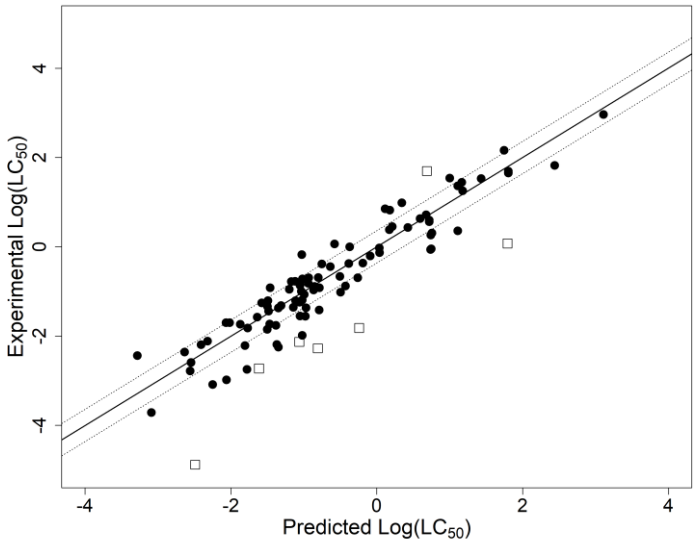

Fig. 3. Experimental vs. predicted 96 h LC_{50} values for fathead minnow. Predictions were obtained by only considering data from the EPAFHM database. The dotted lines represent the intra- laboratory variability factor (2.3). Squares indicate extreme prediction outliers error > 10-fold factor). The logarithmic values refer to concentrations expressed in mM on a linear scale.

The Concordance Correlation Coefficient (CCC) and R^2 that characterize the predictions depicted in Fig. 3 are identical to what described for Fig. 2. Again, these performance indicators highlight the good predictivity of the AW. It is nevertheless interesting to observe that the fact of considering only data obtained within the same laboratory does not improve the predictivity of the AW. This fact suggests that the overall logic of the AW that predict as a function of heterogenous exposure conditions (e.g., flow-through, static, renewal) seems to yield predictions that are precise enough.

QSAR models for acute toxicity in fish are known to display a good predictive performance but it is interesting to observe that more sophisticated QSAR approaches applied to the same fish species are characterized by a comparable (i.e. $0.74 < R^2 < 0.81$) predictive performance in external validation (Jia et al., 2018; Niculescu et al., 2004; Toropova et al., 2012; Wang and Chen, 2020).

3.4 *Analysis of prediction outliers*

The Toolbox offers a convenient and self-contained way to assess modes of actions for aquatic toxicity by means of the "acute aquatic toxicity MOA profiler by OASIS" assign chemicals to different categories according to their acute toxic mode of action. Thanks to theoretical and empiric knowledge the following categories can be identified: Aldehydes, alpha-beta Unsaturated alcohols, Phenols and Anilines, Esters, Narcotic Amines, Basesurface narcotics and a final broad category named "reactive unspecified".

If this profiler (v3.3) is applied to the evaluated chemicals (Supplemental Table S3), it appears that extreme prediction outliers (i.e. prediction error > 10-fold factor) are characterized by a higher proportion of chemicals recognized as having an unspecific reactivity (Tab. 2). The chemicals characterized by this profile among the outliers are: Dibutyl fumarate, 2-Hydroxypropyl acrylate, Dicoumarol, 2,2,2-Trifluoroethanol, Rotenone, 1,1,1,3,3,3-Hexafluoro-2-propanol (for the first evaluation) and Dicoumarol, 2,2,2-Trifluoroethanol and Rotenone (second evaluation).

The null hypothesis of a odds ratio being equal to one can be rejected at a 5% level of significance for the two evaluations (p-values equal to 8 10^{-4} and 0.02 respectively) according to a Fisher's Exact Test for count data.

Table. 2. Two-way contingency tables describing the relationship between extreme prediction outliers and chemicals characterized by an unspecified reactivity.

		Prediction outliers	**Other predictions**
Evaluation 1	*Reactive unspecified*	6	15
	Other profiles	3	98

Evaluation 2	*Reactive unspecified*	3	7
	Other profiles	4	82

These observations would suggest that the predictive performance could be improved by excluding chemicals that are characterized by a MOA profile flagging an unspecific reactivity. Twenty-one and ten chemicals are characterized by such a MOA profile in the framework of the first and second evaluation respectively. If these chemicals are removed, the predictive performance improves. More precisely, R^2 increases to 0.90 and 0.87 for the first and second evaluation respectively. Similarly, CCC increases to 0.94 (first evaluation) and 0.93 (second evaluation).

We deemed that the OASIS profiler provided a reliable and fast option for a preliminary assessment of mode of actions and we did not asses other profilers to avoid the so-called data dredging problem that would have resulted in an increased risk of highlighting false-positive findings (Smith and Ebrahim, 2002). Moreover, the ready availability of the OASIS profiler within the Toolbox renders this assessment of easy application to all the users of the OECD tool.

4. Conclusions

The results herein presented and discussed suggest that the AW for acute fish toxicity is characterized by a predictive performance which is acceptable according to generally adopted quality criteria and comparable to the performance associated with published QSAR models. The fact that the majority of predictions are characterized by a predictive error that is lower than inter-laboratory variability adds support to these findings.

Therefore, the selection of structural analogues performed by the AW and based on the characterization of the mode of action and structural similarity can be considered as particularly appropriate for predictive purposes.

From a regulatory point of view, it should be noted that the AW enables the inspection and documentation, on a case-by-case basis, of the relevance of the toxicological data associated with each structural analogue. For example, a user could eliminate structural analogs associated with data considered to be of insufficient quality and to document this exclusion. In parallel, a user could also choose to treat multiple toxicological data associated with a given chemical by choosing the most conservative value by changing the default option which calculates the arithmetic mean of LC_{50}.

In conclusion, the presented results indicate that, if correctly supervised, the evaluated AW can provide predictions that are reliable and transparent. It is also interesting to note that our findings agree with the results detailed by Burden et al. (Burden et al., 2016) that highlighted the regulatory pertinence and robustness of QSAR predictions for acute fish toxicity. It appears therefore that, if properly used, QSAR approaches can be a valuable tool for providing fit-for-purpose predictions in the framework of regulations on chemical toxicity

Acknowledgements

The financial support from the French Ministry of Ecological Transition is gratefully acknowledged.

Declaration of competing interest

The authors declare the following personal relationships which may be considered as potential competing interests: EM is a member of the OECD QSAR Toolbox management group and PP is the French national coordinator of the OECD Test Guidelines programme.

References

Adelman, I. R., et al., 1976. Fathead Minnows (Pimephales promelas) and Goldfish (Carassius auratus) as Standard Fish in Bioassays and Their Reaction to Potential Reference Toxicants. Journal of the Fisheries Research Board of Canada. 33**,** 209-214.

Barletta, M., Lima, A. R. A., 2019. Systematic Review of Fish Ecology and Anthropogenic Impacts in South American Estuaries: Setting Priorities for Ecosystem Conservation. Frontiers in Marine Science. 6.

Burden, N., et al., 2016. The utility of QSARs in predicting acute fish toxicity of pesticide metabolites: A retrospective validation approach. Regul Toxicol Pharmacol. 80**,** 241-6.

Cappelli, C. I., et al., 2015. Evaluation of QSAR models for predicting the partition coefficient (log P) of chemicals under the REACH regulation. Environ Res. 143**,** 26-32.

Cassano, A., et al., 2014. Evaluation of QSAR models for the prediction of ames genotoxicity: a retrospective exercise on the chemical substances registered under the EU REACH regulation. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev. 32**,** 273-98.

Cherkasov, A., et al., 2014. QSAR modeling: where have you been? Where are you going to? Journal of medicinal chemistry. 57**,** 4977-5010.

Chirico, N., Gramatica, P., 2011. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. J Chem Inf Model. 51**,** 2320-35.

Dimitrov, S. D., et al., 2016. QSAR Toolbox - workflow and major functionalities. SAR QSAR Environ Res. 27**,** 203-219.

EC, 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, The European Parliament and the Council of the European Union. 2006.

Environment-Canada, Guidance document on control of toxicity test precision using reference toxicants 1990.

EPA, Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms. 2002.

EPA, DSSTox (EPAFHM) EPA Fathead Minnow Acute Toxicity. 2020.

FAO, The State of World Fisheries and Aquaculture 2020. Sustainability in action. FAO, Rome, 2020.

Hrovat, M., et al., 2009. Variability of in vivo fish acute toxicity data. Regul Toxicol Pharmacol. 54**,** 294-300.

Jia, Q., et al., 2018. QSAR model for predicting the toxicity of organic compounds to fathead minnow. Environmental Science and Pollution Research. 25**,** 35420-35428.

Johnson, N. L., et al., 1994. Continuous Univariate Distributions. vol. 1. Wiley & Sons.

Könemann, H., 1981. Quantitative structure-activity relationships in fish toxicity studies. Part 1: relationship for 50 industrial pollutants. Toxicology. 19**,** 209-21.

Lind, D., et al., Regional Copper-Nickel Study - Aquatic Toxicology Progress Report -. 1978.

Maechler , M., et al., cluster: Cluster Analysis Basics and Extensions. 2019.

McCarty, L. S., et al., 2018. The regulatory challenge of chemicals in the environment: Toxicity testing, risk assessment, and decision-making models. Regulatory Toxicology and Pharmacology. 99**,** 289-295.

Mombelli, E., 2012. Evaluation of the OECD (Q)SAR Application Toolbox for the profiling of estrogen receptor binding affinities. SAR QSAR Environ Res. 23**,** 37-57.

Netzeva, T., et al., 2008. Review of (Quantitative) Structure–Activity Relationships for Acute Aquatic Toxicity. QSAR & Combinatorial Science. 27**,** 77-90.

Niculescu, S. P., et al., 2004. Using fragment chemistry data mining and probabilistic neural networks in screening chemicals for acute toxicity to the fathead minnow. SAR QSAR Environ Res. 15**,** 293-309.

OECD, 2010. Series on Testing and Assessment No. 126. Short guidance on the threshold approach for acute fish toxicity. Paris.

OECD, 2019. Test No. 203: Fish, Acute Toxicity Test, OECD Guidelines for the Testing of Chemicals, Section 2. Paris.

OECD, The OECD QSAR Toolbox. 2020.

Pickering, Q., Gast, M., 1972. Acute and Chronic Toxicity of Cadmium to the Fathead Minnow (Pimephales promelas). Journal of the Fisheries Research Board of Canada. 29**,** 1099-1106.

R Core Team, A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria, 2019.

Römbke, J., Ahtiainen, J., 2007. The search for the "ideal" soil toxicity test reference substance. Integr Environ Assess Manag. 3**,** 464-6.

Russom, C. L., et al., 1997. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). Environmental Toxicology and Chemistry. 16**,** 948-967.

Schug, H., et al., 2020. Extending the concept of predicting fish acute toxicity in vitro to the intestinal cell line RTgutGC. Altex. 37**,** 37-46.

Smith, G. D., Ebrahim, S., 2002. Data dredging, bias, or confounding. BMJ (Clinical research ed.). 325**,** 1437-1438.

Toropova, A. P., et al., 2012. CORAL: QSAR models for acute toxicity in fathead minnow (Pimephales promelas). J Comput Chem. 33**,** 1218-23.

Walker, J. D., 1988. Relative sensitivity of algae, bacteria, invertebrates, and fish to phenol: Analysis of 234 tests conducted for more than 149 species. Toxicity Assessment. 3**,** 415-447.

Wang, Y., Chen, X., 2020. A joint optimization QSAR model of fathead minnow acute toxicity based on a radial basis function neural network and its consensus modeling. RSC Advances. 10**,** 21292-21308.

Wehrens, R., 2011. Chemometrics with R. Springer-Verlag, Berlin Heidelberg.

Yap, C. W., 2011. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem. 32**,** 1466-74.

Yordanova, D., et al., 2019. Automated and standardized workflows in the OECD QSAR Toolbox. Computational Toxicology. 10**,** 89-104.