Journal of Visualized Experiments

In Silico Modeling Method for Computational Aquatic Toxicology of Endocrine Disruptors: A Software-Based Approach Using QSAR Toolbox

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE60054R1
Full Title:	In Silico Modeling Method for Computational Aquatic Toxicology of Endocrine Disruptors: A Software-Based Approach Using QSAR Toolbox
Keywords:	OECD QSAR Toolbox automated workflow; quantitative structure-activity relationship (QSAR); endocrine disrupting chemical; aquatic vertebrate, acute toxicity; computational ecotoxicology
Corresponding Author:	Baeckkyoung Sung GERMANY
Corresponding Author's Institution:	
Corresponding Author E-Mail:	sung@kist-europe.de
Order of Authors:	Marie-Léonie Bohlen
	Hyun Pyo Jeon
	Young Jun Kim
	Baeckkyoung Sung
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Open Access (US$4,200)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Saarbrücken, Saarland, Germany

Dear Editor of *JoVE*,

We would like to submit our revision, titled as "*In Silico* Modeling Method for Computational Aquatic Toxicology of Endocrine Disruptors: A Software-Based Approach Using QSAR Toolbox", for consideration of publication in *Journal of Visualized Experiments*.

We have extensively revised the manuscript according to the comments of reviewers and did our best to answer their questions. Thanks to the valuable advice from the reviewers, we believe that our manuscript has been prominently improved compared to its original version. We hope that this revision could have fulfilled the criteria for being accepted.

Sincerely yours,

Baeckkyoung Sung, Ph.D.

KIST Europe Forschungsgesellschaft mbH

66123 Saarbrücken, Germany

E-mail: sung@kist-europe.de

Tel: +49 (0)681-9382-379

TITLE:

Using the QSAR Toolbox as an In Silico Modeling Method for Computational Aquatic Toxicology of Endocrine Disruptors

AUTHORS AND AFFILIATIONS:

Marie-Léonie Bohlen[1], Hyun Pyo Jeon[1], Young Jun Kim[1], Baeckkyoung Sung[1]

[1]KIST Europe Forschungsgesellschaft mbH, Saarbrücken, Germany

Corresponding Author:

Baeckkyoung Sung (sung@kist-europe.de)
Tel: +49 (0)681-9382-379

Email Addresses of Co-authors:

Marie-Léonie Bohlen (ml.bohlen@kist-europe.de)
Hyun Pyo Jeon (hpjeon@kist-europe.de)
Young Jun Kim (youngjunkim@kist-europe.de)

KEYWORDS:

OECD QSAR Toolbox, automated workflow, quantitative structure-activity relationship, QSAR, endocrine disrupting chemical, aquatic vertebrate, acute toxicity, computational ecotoxicology

SUMMARY:

Quantitative structure-activity relationship (QSAR) modeling is a representative bioinformatics-assisted method in toxicological screening. This protocol demonstrates how to computationally assess the risks of endocrine disruptors (EDs) in aquatic environments. Utilizing the OECD QSAR Toolbox, the protocol implements an in silico assay for analyzing toxicity of EDs in fish.

ABSTRACT:

Computational analyses of toxicological processes enables high-throughput screening of chemical substances and prediction of their endpoints in biological systems. In particular, quantitative structure-activity relationship (QSAR) models have been increasingly applied to assess the environmental effects of a plethora of toxic materials. In recent years, some more highlighted types of toxicants are endocrine disruptors (EDs, which are chemicals that can interfere with any hormone-related metabolism). Because EDs may significantly affect animal development and reproduction, rapidly predicting the adverse effects of EDs using in silico techniques is required. This study presents an in silico method to generate prediction data on the effects of representative EDs in aquatic vertebrates, particularly fish species. The protocol describes an example utilizing the automated workflow of the QSAR Toolbox software developed by the Organization for Economic Co-operation and Development (OECD) to enable acute ecotoxicity predictions of EDs. As a result, the following are determined: (1) calculation of the numerical correlations between the concentration for 50% of lethality (LC_{50}) and octanol-water partition coefficient (K_{ow}), (2) output performances in which the LC_{50} values determined

in experiments are compared to those generated by computations, and (3) the dependence of estrogen receptor binding affinity on the relationship between K_{ow} and LC_{50}.

INTRODUCTION:

New developments in informatics and computational technology have empowered the biological sciences with quantitative methodologies that offer high precision and reliability[1]. In particular, algorithms used in molecular taxonomy and property classification have resulted in quantitative structure-activity relationship (QSAR) models[2]. These models automatically correlate the chemical structures and biological activities of a given chemical database and implement rapid in silico screening of a wide range of chemical substrates according to their medicinal or toxicological actions[3]. QSAR tools can produce predictive toxicity profiles as a function of feature vectors of molecular descriptors (i.e., physicochemical parameters) of chemicals of interest to numerically create categorical endpoints[4]. Usually, each quantitative endpoint is displayed as a 2D scatterplot vs. changes in descriptor values. A QSAR model is then generated using (multiple) linear regression analyses. Once a dataset has been fully exploited to construct a QSAR model (called the training set), then the model is statistically validated by predicting the endpoints of a group of chemicals not included in the training set (called the test set). The model can then be used to predict the biological activities of untested compounds[3].

Among many harmful chemicals, endocrine disruptors (EDs) have been highlighted as a group of toxicants that may interfere in numerous hormone-related metabolisms in mammals, amphibians, and fish[5,6]. EDs are known to induce a variety of adverse effects, such as cancers and malformations, by blocking or altering normal hormonal pathways or activating abnormal hormone synthesis/degradation signals. As a consequence, these hormone-mimicking chemicals can perturb endocrine systems such that biological development and reproduction of wildlife animal populations are hampered. In particular, the ecotoxicological effects of EDs have been extensively investigated in aquatic vertebrates, which have nearly identical hormone receptor structures to those of mammals, including humans. Because all hormonal actions occur at low doses in vivo, predicting the potential toxicities of ED candidates using rapid in silico screening is critical to public and environmental health.

QSAR models based on the toxicology of EDs have been conducted utilizing both 2D and 3D descriptors (known as 2D and 3D QSAR, respectively), which reveal the ED ligand binding affinities of estrogen, androgen, and progesterone receptors[7]. Despite the high-precision advantages of 3D QSAR, in which conformational and electrostatic interactions are considered, 2D QSAR retains its own robustness in direct mathematical algorithms, rapid calculations, and extremely low computational loads. In addition, 2D-QSAR models are flexible for use in a wide range of applications while achieving relatively accurate prediction performance.

The OECD QSAR Toolbox is currently one of the most utilized computer software tools, providing freely available and pre-built QSAR models[8,9]. Its profiler uses 2D descriptor databases. Since the release of the first version in 2008, the software has been applied in the fields of chemical and biological industries, public health, and environmental safety for full or

partial analysis of the potential risks of natural and synthetic compounds, with special interests in carcinogenesis[10-12], mutagenicity[13-15], and developmental toxicity[16]. The application to aquatic toxicology has also been demonstrated, with focus on bioaccumulation and biotransformation[17].

The QSAR Toolbox has been proven useful in predicting the short-term toxicity of a broad range of chemicals[17], as well as the estrogen receptor (ER) binding affinities of EDs[18]. However, the acute ecotoxicities of EDs in aquatic vertebrates has not been analyzed using the QSAR Toolbox. In this study, a typical and facile protocol is presented to perform QSAR modeling on the acute adverse effects of EDs with a focus in fish species. The study shows that the QSAR Toolbox is a highly accessible software for calculating and predicting the lethality/mortality of aquatic vertebrates for some representative EDs. Statistical treatment methods for the derived in silico datasets are presented. **Figure 1** shows the overall scheme for the general operation of the QSAR Toolbox. The workflow shown in **Figure 2** provides straightforward instructions on how to operate the in silico assay to predict acute ecotoxicity of target substances such as endocrine disrupting chemicals.

PROTOCOL:

1. Equipment

1.1. Software: use OECD QSAR Toolbox 4.0 or newer (free download from <https://qsartoolbox.org/download/?) and data analysis software.

1.2. Computer: for the OECD QSAR Toolbox, use: (i) system type: 64 bit, Windows 7 or newer; (ii) processor: I5 at 2.4 GHz, or a faster processor or equivalent AMD CPU; (iii) installed memory (RAM): 6 GB; (iv) hard disk drive (HDD): 20 GB of free hard drive space (OECD QSAR Toolbox 4.3 Release Notes: <https://qsartoolbox.org/file/2019/02/Toolbox-4.3-Release-Notes-1.pdf>).

2. Procedure

2.1. OECD QSAR Toolbox

NOTE: The QSAR Toolbox operates in six consecutive flow modules starting from **Input** and followed by **Profiling, Data, Category Definition, Data Gap Filling,** then **Report,** located at the top of the program interface.

2.1.1. Explore the aforementioned six stages through six toolbar icons by left-clicking. First, look over the stages of **Input, Data Gap Filling,** and **Report** that are necessary to perform the automated workflow “Ecotoxicological endpoint” and to document its results.

2.1.2. Take a short look over optional stages **Profiling** and **Data.** The **Profiling** stage provides an initial insight into the target substance’s (eco)toxicity potential and environmental fate

characteristics. Optional **Data** stage enables searching for available experimental data related to the target substance.

2.2. Input

2.2.1. Upon starting the QSAR Toolbox, the user begins at the **Input** toolbox stage by default. The QSAR Toolbox creates a working file named "Document 1" automatically, which is displayed in the stage option panel on the left of the program interface. Rename the file, if desired, by right-clicking the working file.

2.2.2. Click on the **CAS#** button in the actions toolbar, enter the chemical abstract service (CAS) number of the target substance in the available text field, and click **Search.** The tool will then searche for the target substance by CAS number.

2.2.3. If required, choose other search options that are available in the action toolbar such as searching by substance name or simplified molecular-input line-entry system (SMILES) code. SMILES can be entered as 2D non-stereochemical or 3D stereochemical containing forms. Click **Name** or **Structure,** respectively. Use the **Structure** tool to draw the target substance.

2.2.4. The search tool displays the search results through database records in a pop-up window. Choose the record reporting a "high" CAS-SMILES relation (**CS Relation** field) if multiple records are retrieved for the target substance by checking the box on the left of the record. Click **OK**.

NOTE: Proceeding from this point is possible only if the retrieved record contains a SMILES code, as the SMILES code (2D non-stereochemical containing form) is the basis for computation.

2.2.5. Batch mode: to perform the in silico assay for multiple target substances, write a simple substance list in a text editor in which each CAS number is listed in a single row (**Supplementary Figure S3**). Save the text file with an appropriate name and extension .txt on the computer.

2.2.6. Batch mode: click **Data.** Then, go to **Databases** in the stage option panel on the left of the program interface. Make sure databases that are listed under **Ecotoxicological Information** are checked**.**

2.2.7. Batch mode: click **Input.** Select **Query** from the actions toolbar. Accept the settings set in step 2.2.6 by clicking **Yes** in the dialog window.

2.2.8. Batch mode: choose the **CAS** tab. Upload the substance list saved as text file through **Load list** from your computer.

2.2.9. Batch mode: there are two **Add** buttons available; click the **Add** button on the bottom of the pop-up menu and then click **Execute.** The QSAR Toolbox will display a message on the number of substances that have been retrieved for the search.

NOTE: Some substances of the loaded list may not be found by the search tool or that several entries may be available for one CAS number. It is not possible to delete substances from the retrieved set of substances.

2.3. Profiling

NOTE: The following section is optional. If this is not required, skip to section 2.5.

2.3.1. Click on the toolbox stage button **Profiling.** Go to **Profiling methods** in the stage option panel on the left of the program interface.

2.3.2. Click **Unselect All.** Check all profilers listed under **Predefined** and those related to aquatic toxicity listed under **Endpoint specific** such as "Acute aquatic toxicity classification by Verhaar (Modified)."

2.3.3. Finish the selection. Then click on the **Apply** button in the Actions toolbar.

NOTE: The QSAR Toolbox provides recommendations on a set of profilers. These are highlighted in green (suitable) and orange (plausible) when left-clicking the data matrix field next to the endpoint of interest. Available endpoints are listed in the endpoint tree next to the stage option panel. The profiler **Substance type** will indicate whether the target substance is a "discrete chemical." The information is displayed in the expanded endpoint tree "Profile", "Predefined", and "Substance type". Only if the target substance is a discrete chemical can the automated workflow run successfully. "Acute aquatic toxicity classification by Verhaar (modified)" provides a first estimate of the acute aquatic toxicity mechanism of the target substance[19,20]. The information is displayed in the expanded endpoint tree "Profile", "Endpoint Specific", and "Acute aquatic toxicity classification by Verhaar (modified)". Five classes are available: (class 1) inert chemicals (baseline toxicity); (class 2) less inert chemicals; (class 3) reactive chemicals; (class 4) specifically acting chemicals; and (class 5) for chemicals not possible to classify.

2.3.4. Click **Parameter** in the endpoint tree to run integrated 2D and 3D QSAR models available in the QSAR Toolbox, if desired. Click **Calculate/extract all parameters for all chemicals** in the pop-up menu.

2.3.5. 2D and 3D QSAR models compiled in **Parameter** provide numeric values. Use "Profiling methods" for qualitative information (see step 2.3.1).

2.4. Data

NOTE: This section is optional. If it is not required, skip to section 2.5.

2.4.1. Click on the toolbox stage button **Data.** Then, click **Gather** from the Actions toolbar.

2.4.2. Select **All endpoints** to gather all experimental data, then **Choose** to gather endpoint specific experimental data. As an example, if aquatic toxicity is the user's focus, click **Choose** > **Ecotoxicological Information** > **Aquatic toxicity** > **OK.**

NOTE: Choosing to gather experimental data for all endpoints may lead to extended processing time. The user can adapt the hierarchy of the endpoint tree to the specific purpose. This changes the manner in which data are displayed.

2.4.3. If desired, right-click the endpoint of interest in the endpoint tree area. Choose **Set tree hierarchy** in the pop-up menu. Organize the endpoint tree in the preferred manner using the available terms and arrows and click **OK.**

2.4.4. If desired, export the gathered data as an Excel file. Right-click on the endpoint of interest and choose **Export Data matrix** in the pop-up menu.

2.4.5. A "Matrix export" wizard opens and enables adding other endpoints to the export list. Finish the selection, click **Export,** and save the file on the computer.

NOTE: Exporting data from all databases is not possible. For example, data retrieved from the database "ECHA CHEM" cannot be saved.

2.5. Data gap filling

2.5.1. Click on the toolbox stage button **Data gap filling.** Then, click **Automated** in the Actions toolbar.

2.5.2. Select **Ecotoxicological Endpoint** > **Fish, LC50** (lethal concentration, 50%) at 96 h for *Pimephales promelas* (mortality). Click **OK.** A "Workflow Controller" will appear, and processing will take up to several minutes, especially in batch mode.

NOTE: The QSAR Toolbox automatically applies a defined set of profilers when searching for suitable substances with available experimental data for the prediction. The experimental data [e.g., effect concentrations 96 h LC_{50} (*P. promelas*) or 96 h EC_{50} (*P. promelas*, mortality)] are used to generate the prediction for the target substance by either linear approximation or nearest neighbor method. Note that the methods of linear approximation and nearest neighbor are referred to as trend analysis (labeled as "T") and read-across (labeled as "R"), respectively.

2.5.3. The user will receive a message if the prediction is performed successfully. Click **OK** and close the workflow controller indicating "Finished workflow" by clicking **x** in the upper right corner.

2.5.4. Batch mode: upon starting the automated workflow, the user will be asked to specify the range of substances over which to execute the workflow. Accept the full range of substances selected by default in the dialog window by clicking **OK.**

2.5.5. Batch mode: the user will not receive a message indicating whether a prediction was run successfully or unsuccessfully. Close the workflow controller indicating “Finished workflow” at the end of the batch processing by clicking **x** in the upper right corner.

2.6. Report

2.6.1. If a prediction was successfully executed, click on the toolbox stage button **Report.**

NOTE: No reports can be generated in batch mode.

2.6.2. Scroll down and find the prediction value in matrix field located in a yellow highlighted row next to endpoint “96-h.” The predicted value is labelled with “T” or “R.” Activate this specific data matrix field by left-clicking it.

2.6.3. Click **Prediction** in the Actions toolbar. Customize the report content and appearance in the pop-up wizard. Three types of reports are available: (i) prediction, (ii) category, and (iii) data matrix.

2.6.4. The wizard allows the user to fill in the author’s name and contact details. Write a short summary, provide a detailed explanation of the mechanistic interpretation, or provide justification for the adequacy of the prediction.

2.6.5. Include additional information related to the executed prediction, if desired. The extent of additional information depends on the user.

2.6.6. Go through the wizard by clicking **Next.** Finally, click **Create report** and save the prediction and category reports as PDF files and the data matrix as an Excel spreadsheet on the computer.

2.6.7. Find additional details on the functionalities of the QSAR Toolbox and automated workflows in the application manual for the OECD QSAR Toolbox v.4 (F1 help on the keyboard). Details on the algorithms and rationale behind the automated workflow are described by Dimitrov et al.[8] and Yordanova et al. [9].

3. Application

3.1. If using the predicted effect concentration (i.e., 96-h LC_{50} of *P. promelas*) in the environmental risk assessment, use the lower limit of the 95% confidence interval. Find the data on the first page of the saved prediction report (PDF) in “Prediction summary”, “Predicted value: <mean> (from <lower_limit> to <upper_limit>).”

NOTE: The notes given here are based on results of the comparison between predicted and experimental data for a set of target substances reported in this study. Selecting the lower end

of the 95% confidence range will increase the likelihood that the predicted effect concentration will not underestimate the real toxicity of the substance (see the representative results). The predicted effective concentration of the lower limit of the 95% confidence interval will therefore present a safer basis for risk assessment.

REPRESENTATIVE RESULTS:

The example described in this study was implemented for quantitative analysis and prediction of acute toxicities of selected EDs in fish. When the predicted data points were plotted versus experimental data points as a log-log scale, a positive correlation between both was found for all fish and a representative species, namely, *Pimephales promelas* (fathead minnow; **Figure 3**). In both cases, the slope of the linear regression appeared to be comparable (predicted LC_{50}/experimental LC_{50} = 0.611 and 0.602 for all fish and *P. promelas*, respectively). Because of the limited amount of experimental data, the number of available values from experimental observation was usually smaller than that from computational prediction. Application of the tolerance factor as 5-fold for the computational capability[21] resulted in 94% (34/36) and 96% (26/27) of the protective prediction for all fish and *P. promelas*, respectively. Based on this prediction, 3',5,7-trihydroxy-4',6-dimethoxyisoflavone and 1,4-benzenediol appeared to exhibit calculated LC_{50} values greater than the tolerance limit.

To enable safety assessment at the highest reliability, further computational analysis was performed by plotting the predicted lower limit of the 95% confidence interval of LC_{50} (instead of the mean values used in **Figure 3**) versus the experimentally derived values (**Figure 4**). In this evaluation with an elevated safety threshold, 92% (33/36) of the total tested endocrine disrupting compounds were shown to fall into the protective range when compared to the experimentally derived values except for: 3',5,7-trihydroxy-4',6-dimethoxyisoflavone; 1,4-benzenediol; and 4-hexylphenol.

Based on assessments of the entire species available from the database, values for the predicted and experimental 96-h $\log_{10}LC_{50}$ exhibited linearity with the $\log_{10}K_{OW}$ values in the domain between -1 and 7, indicating a hyperbolic correlation between LC_{50} and K_{OW}. An overall trend existed whereby the LC_{50} decreased for higher K_{OW} values of EDs for the data obtained from both computational predictions and experiments, suggesting increasing acute toxicity in fish species for EDs with higher hydrophobicity (**Supplementary Figure S1**).

By the rule-based ER profiler embedded in the OECD QSAR Toolbox, the ER binding affinities of the EDs were categorized as non-binding as well as weak, moderate, strong, and very strong binders, in order of increasing binding affinity[18]. Accordingly, the statistical distribution of $\log_{10}K_{ow}$ could be displayed as a qualitative classification of ER binding affinity (**Supplementary Figure S2**). Overall, the changes in K_{ow} distribution ranges and their mean levels appeared to not have a defined tendency. Similarly, the distributions of predicted and experimental LC_{50} were shown as the extent of ER binding affinity (**Figure 5**). In this case, mean levels of predicted LC_{50} for ER binders were higher than those of non-binders. By contrast, for the experimental LC_{50}, the mean levels of non- and weak binders were higher than those of stronger ER binders.

FIGURE AND TABLE LEGENDS:

Figure 1: Basic scheme of the general workflow of the OECD QSAR Toolbox.

Figure 2: Workflow. Shown is the workflow conceptualizing the modules and sequences applied to predict the acute toxicities of endocrine disruptors (EDs) in fish using the OECD QSAR Toolbox.

Figure 3: Predicted vs. experimental 96-h LC_{50} of EDs in Table 1 for all fish (blue diamonds, n = 36) and a selected species *P. promelas* (cyan diamonds, n = 27). For the predicted LC_{50}, the average ("AVE") values are displayed. The dashed lines represent linear regressions for the two groups: for all fish (light blue), predicted LC_{50}^{AVE} = 0.611 x (experimental LC_{50}) + 0.277 (adjusted r^2 = 0.408); and for *P. promelas* (light cyan), predicted LC_{50}^{AVE} = 0.602 x (experimental LC_{50}) + 0.385 (adjusted r^2 = 0.441). The solid diagonal line shows unity in which the predicted and experimental values are equal[21]. The dotted gray line shows the 5-fold tolerance limit of the computational capability[19]. Outliers: 3',5,7-trihydroxy-4',6-dimethoxyisoflavone (*) and 1,4-benzenediol (**).

Figure 4: Predicted (lower limit of 95% confidence interval, "low-95%") vs. experimental 96-h LC_{50} of EDs in Table 1 for all fish (n = 36). The dashed line represents the linear regression: predicted LC_{50}^{AVE} = 0.470 x (experimental LC_{50}) - 0.312, where adjusted r^2 = 0.193. The solid diagonal line indicates unity where the predicted and experimental values are equal to each other[19]. Outliers: 3',5,7-trihydroxy-4',6-dimethoxyisoflavone (*), 1,4-benzenediol (**), and 4-hexylphenol (***).

Figure 5: Distributions of predicted (solid boxes, n = 8–20 for each category) and experimental (dashed boxes; n = 3–16 for each category) 96-h LC_{50} depending on ER binding affinity of EDs in Table 1 for all fish. A box plot represents: (**A**) mean (small square with a horizontal bar), (**B**) 1st and 3rd quartiles (lower and upper ends of the box, respectively), (**C**) median (horizontal segment inside the box), (**D**) 5th and 95th percentile (lower and upper error bars, respectively), (**E**) 1st and 99th percentile (lower and upper x, respectively), and (**F**) minimum and maximum (lower and upper "-", respectively).

Table 1: List of evaluated endocrine disrupting chemicals. Average mean (AVE) and lower 95% confidence interval (CI) effective concentrations (95-h LC_{50}, *Pimephales promelas*) as well as Estrogen Receptor Binding were predicted with the QSAR Toolbox version 4.3 Automated Workflow. $Log_{10}K_{ow}$ was retrieved via QSAR Toolbox version 4.3 from KOWWIN v1.68, 2000, U.S. Environmental Protection Agency. Experimental $log_{10}K_{ow}$ values were preferred over predicted values. The target substance list was compiled from previously reported lists of EDs[22-24].

DISCUSSION:

The versatility of the OECD QSAR Toolbox as analytic software for ecotoxicology is shown here with specific interest in the adverse effects of endocrine disrupting chemicals on aquatic vertebrates. In addition, a simple and standard protocol was demonstrated for predicting acute toxicity (96-h LC_{50}) of 74 representative EDs (**Table 1**) for fish species. This was achieved by applying category building, data gap filling, and ER profiling modules embedded in the QSAR Toolbox (**Figure 1, Figure 2**).

The linear correlation between $\log_{10}LC_{50}$ and $\log_{10}K_{OW}$ with a negative slope (as shown in **Supplementary Figure S1**) has long been known as a standard quantitative relationship in QSAR analyses[25], where higher toxicity is shown the more hydrophobic a given chemical is. As can be seen from a simple calculation, the general mathematical relation that includes **Equation 1** and **Equation 2** is a converted expression from the following power function[26]:

$$log(LC_{50}) = a' - b \cdot log(K_{ow}) \qquad (1)$$

$$LC_{50} = a \cdot {K_{ow}}^{-b}, \text{ where } a' = log(a) \qquad (2)$$

From the plot of (3), characterizing an intermediate range of K_{OW}[26] may be possible by adjusting the parameters a and b, where a certain variation in hydrophobicity (or hydrophilicity) does not significantly change the endpoint of acute toxicity.

Comparative analyses between the computational predictions and experimental observations on the LC_{50}, as shown in **Figure 3** and **Figure 4**, have been typically reported in studies of QSAR for various aquatic toxicants, including technical nonionic surfactants[27], triazole fungicides[28], and pesticide metabolites[21]. This type of retrospective validation provides information on how far a given QSAR tool can reach in terms of comparative performance to experimental results. In this study of acute toxicity in fish, the QSAR Toolbox was proven to provide protective predictions for over 90% of tested EDs in all fish and in a single species, *Pimephales promelas*.

Further identifying the three outlier chemicals in **Figure 3** and **Figure 4**, which showed higher predicted LC_{50} on average and at a minimum, respectively, is required. First, the 3',5,7-trihydroxy-4',6-dimethoxyisoflavone is a type of flavonoid (more specifically, an isoflavone), which is considered to be generally safe and used in herbal pharmaceuticals; however, it still has estrogen-related concerns[29] and may cause acute toxicity probably through oxidative phosphorylation uncoupling[30]. Next, the 1,4-benzenediol, called hydroquinone, is a phenolic compound that can trigger a non-specific and cytotoxic immune response in fish[31]. Finally, the 4-hexylphenol has been known to exhibit sufficient positive estrogenic activity to be classified as an ED[32]. It has been well-studied that the main reason of the acute toxicity of hydroquinone is the reduction-oxidation (redox) cycling. The hydroquinone is oxidized to benzoquinone and reduced back to semi-quinone or hydroquinone repeatedly, with depleting cofactors and generating reactive oxygen species[33]. The other two chemicals may require deeper investigations to reveal their mechanisms of action in acute ecotoxicity using molecular docking approaches such as that used by Panche et al.[34], which cannot be covered by the QSAR Toolbox.

EDs interfere with the endocrine system mainly through physicochemical interactions with steroid receptors such as estrogen and androgen receptors, which are of considerable interest in QSAR modeling studies[35]. Considering this, the QSAR Toolbox is robust in terms of facile and rapid classification of ER binding affinities for a set of chemicals based only on the 2D descriptors of molecular structures. When this ER profiler system was applied to our list of EDs, no clear correlation was found between ER binding affinity and hydrophobicity (**Supplementary Figure S2**). This result may be explained by the fact that the formation of a steroid-receptor complex is not a direct consequence of a hydrophobic bonding contribution but should be accompanied by a conformational change in the active-site receptor structure[36]. The receptor binding can be also due to hydrogen-bonding and π-stacking.

Additionally, the position of each chemical group on the molecule may affect the receptor binding, even if the hydrophobicity and number of hydrogen-bond acceptors-donors remain the same. Second, the ER profiler produced contrary trends between predicted and experimental LC_{50} mean levels with increasing ER binding affinity (**Figure 5**). This may be because the lethality of parents in an acute toxicity test are not due to ER binding but rather to narcosis in most cases, or to redox cycling in the case of hydroquinone. For example, more extensive analysis, including the chronic toxicity, is required for a larger set of EDs to define predictive limitations of the current version of the QSAR Toolbox.

This preliminary research may also have public health implications because steroids (androgens, estrogens, progestines, and corticoids) and their receptors exhibit similar or even identical macromolecular structures across vertebrates[5]. These types of analogous endocrine signaling systems may operate using a common mechanism in key events of EDs[5]. Nevertheless, additional and complementary methodologies are required to illuminate this vast and complex aspect [for example, by performing computational modeling of absorption, distribution, metabolism, and excretion (ADME), and/or adverse outcome pathway (AOP)][38]. Furthermore, because most of the scientific and public concerns raised about the adverse effects of EDs are related to their chronic toxicities, improving the databases and algorithms in the QSAR Toolbox and producing reliable long-term ecotoxicology predictions for EDs are both necessary.

This paper demonstrates the application of QSAR Toolbox to compare ecotoxicological LC_{50} values for fish with $\log_{10}K_{ow}$ values of EDs. Throughout the protocol, it results in weak relationships between the two parameters, as it has been revealed by previous studies (e.g., Kim et al.[39]) that $\log_{10}K_{ow}$ is not a good direct predictor of aquatic LC_{50}. In spite of this limitation, this protocol provides a general review or “vignette” to describe how to use the dashboard for a given purpose, since it is a valid application to use the QSAR Toolbox for investigating correlations between LC_{50} (or ER binding affinity) and $\log_{10}K_{ow}$. Nevertheless, it should be noted that (1) illuminating the link between estrogen receptor binding and chronic toxicity, rather than acute toxicity (lethality), is more relevant so that clearer correlations may be found, and (2) the androgen receptor, together with that of estrogen, also plays a critical role in reproductive toxicity. Therefore, it is required for the future version of the QSAR Toolbox to improve the prediction functions in light of those two points.

ACKNOWLEDGMENTS:

This research was supported by the National Research Council of Science & Technology (NST) grant by the South Korean government (MSIP) (No. CAP-17-01-KIST Europe) and Project 11911.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1 Najarian, K., Najarian, S., Gharibzadeh, S., Eichelberger, C. N. *Systems Biology and Bioinformatics: A Computational Approach.* CRC Press. Boca Raton, FL, USA (2009).

2 Fujita, T., Iwasa, J., Hansch, C. A new substituent constant, π, derived from partition coefficient. *Journal of the American Chemical Society.* **86**, 5175-5180 (1964).

3 Roy, K., Kar, S., Das, R. N. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment.* Academic Press. Cambridge, MA, USA (2015).

4 Raies, A. B., Bajic, V. B. In silico toxicology: computational methods for the prediction of chemical toxicity. *WIREs Computational Molecular Science.* **6**, 147-172 (2016).

5 Hayes, T. B. Welcome to the revolution: integrative biology and assessing the impact of endocrine disruptors on environmental and public health. *Integrative Compuational Biology.* **45**, 321-329 (2005).

6 Schug, T. T. et al. Minireview: endocrine disruptors: past lessons and future directions. *Molecular Endocrinology.* **30**, 833-847 (2016).

7 Devillers, J., Marchand-Geneste, N., Carpy, A., Porcher, J. M. SAR and QSAR modeling of endocrine disruptors. *SAR QSAR Environmental Research.* **17**, 393-412 (2006).

8 Dimitrov, S. D. et al. QSAR Toolbox - workflow and major functionalities. *SAR QSAR Environmental Research.* **27**, 203-219 (2016).

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

10 Mombelli, E., Devillers, J. Evaluation of the OECD (Q)SAR Application Toolbox and Toxtree for predicting and profiling the carcinogenic potential of chemicals. *SAR QSAR Environmental Research.* **21**, 731-752 (2010).

11 Devillers, J., Mombelli, E., Samsera, R. Structural alerts for estimating the carcinogenicity of pesticides and biocides. *SAR QSAR Environmental Research.* **22**, 89-106 (2011).

12 Li, C. et al. Identifying unknown by-products in drinking water using comprehensive two-dimensional gas chromatography-quadrupole mass spectrometry and in silico toxicity assessment. *Chemosphere.* **163**, 535-543 (2016).

13 Devillers, J., Mombelli, E. Evaluation of the OECD QSAR Application Toolbox and Toxtree for estimating the mutagenicity of chemicals. Part 1. Aromatic amines. *SAR QSAR Environmental Research.* **21**, 753-769 (2010a).

14 Devillers, J., Mombelli, E. Evaluation of the OECD QSAR Application Toolbox and Toxtree for estimating the mutagenicity of chemicals. Part 2. α-β unsaturated aliphatic aldehydes. *SAR QSAR Environmental Research.* **21**, 771-783 (2010b).

15 Kulkarni, S. A., Barton-Maclaren, T. S. Performance of (Q)SAR models for predicting Ames mutagenicity of aryl azo and benzidine based compounds. *Journal of*

Environmental Science and Health Part C Environmental Carcinogenesis & Ecotoxicology Reviews. **32**, 46-82 (2014).

16 Craig, E. A., Wang, N. C., Zhao, Q. J. Using quantitative structure-activity relationship modeling to quantitatively predict the developmental toxicity of halogenated azole compounds. *Journal of Applied Toxicology.* **34**, 787-794 (2014).

17 Tebby, C., Mombelli, E., Pandard, P., Péry, A. R. Exploring an ecotoxicity database with the OECD (Q)SAR Toolbox and DRAGON descriptors in order to prioritise testing on algae, daphnids, and fish. *Science of the Total Environment.* **409**, 3334-3343 (2011).

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

19 Verhaar, H. J. M., van Leeuwen, C. J., Hermens, J. L. M. Classifying environmental pollutants. 1: structure-activity relationships for prediction of aquatic toxicology. *Chemosphere.* **25**, 471-491 (1992).

20 Enoch, S. J., Hewitt, M., Cronin, M. T. D., Azam, S., Madden, J. C. Classification of chemicals according to mechanism of aquatic toxicity: an evaluation of the implementation of the Verhaar scheme in Toxtree. *Chemosphere.* **73**, 243-248 (2008).

21 Burden, N., Maynard, S. K., Weltje, L., Wheeler, J. R. The utility of QSARs in predicting acute fish toxicity of pesticide metabolites: a retrospective validation approach. *Regulatory Toxicology and Pharmacology.* **80**, 241-246 (2016).

22 Nendza, M. et al. Screening for potential endocrine disruptors in fish: evidence from structural alerts and in vitro and in vivo toxicological assays. *Environmental Sciences Europe.* **28**, 26 (2016).

23 Roncaglioni, A., Piclin, N., Pintore, M., Benfenati, E. Binary classification models for endocrine disrupter effects mediated through the estrogen receptor. *SAR QSAR Environmental Research.* **19**, 697-733 (2008).

24 Sosnovcová, J., Rucki, M., Bendová, H. Estrogen receptor binding affinity of food contact material components estimated by QSAR. *Central European Journal of Public Health.* **24**, 241-244 (2016).

25 Walker, J. D., Dearden, J. C., Schultz, T. W., Jaworska, J., Comber, M. H. I. QSARs for New Practitioners. In: Walker, J. D. (ed.) *QSARs for Pollution Prevention, Toxicity Screening, Risk Assessment, and Web Applications*. SETAC Press. Pensacola, FL, USA (2003).

26 Sánchez-Bayo, F. From simple toxicological models to prediction of toxic effects in time. *Ecotoxicology.* **18**, 343-354 (2009).

27 Sjöström, M., Lindgren, Å., Uppgård, L-L. Joint Multivariate Quantitative Structure-Property and Structure-Activity Relationships for a Series of Technical Nonionic Surfactants. In: Chen, F., Schüürmann, G. (eds.) *Quantitative Structure-Activity Relationships in Environmental Sciences-VII.* SETAC Press. Pensacola, FL, USA (1997).

28 Ding, F., Guo, J., Song, W., Hu, W., Li, Z. Comparative quantitative structure-activity relationship (QSAR) study on acute toxicity of triazole fungicides to zebrafish. *Chemistry Ecology.* **27**, 359-368 (2011).

29 Galati, G., O'Brien, P. J. Potential toxicity of flavonoids and other dietary phenolics: significance for their chemopreventive and anticancer properties. *Free Radical Biology in Medicine.* **37**, 287-303 (2004).

30 Russom, C. L., Bradbury, S. P., Broderius, S. J. Predicting modes of action from chemical structure: acute toxicity in the fathead minnow (Pimephales promelas). *Environmental Toxicology and Chemistry*. **16**, 948-967 (1997).
31 Taysse, L., Troutaud, D., Khan, N. A., Deschaux, P. Structure-activity relationship of phenolic compounds (phenol, pyrocatechol and hydroquinone) on natural lymphocytotoxicity of carp (*Cyprinus carpio*). *Toxicology.* **98**, 207-214 (1995).
32 Nishihara, T. et al. Estrogenic activities of 517 chemicals by yeast two-hybrid assay. *Journal of Health Science.* **46**, 282-298 (2000).
33 Bolton, J. L., Trush, M. A., Penning, T. M., Dryhurst, G., Monks, T. J. Role of quinones in toxicology. *Chemical Research in Toxicology.* **13**, 135-160 (2000).
34 Panche, A. N., Diwan, A. D., Chandra, S. R. Flavonoids: an overview. *Journal of Nutritional Science.* **5**, e47 (2016).
35 Li, J., Gramatica, P. QSAR classification of estrogen receptor binders and pre-screening of potential pleiotropic EDCs. *SAR QSAR Environmental Research.* **21**, 657-669 (2010).
36 Bohl, M. *Molecular Structure and Biological Activity of Steroids*. CRC Press. Boca Raton, FL, USA (2017).
37 Kaminuma, T., Takai-Igarashi, T., Nakano, T., Nakata, K. Modeling of signaling pathways for endocrine disruptors. *BioSystems.* **55**, 23-31 (2000).
38 Lillicrap, A. et al. Alternative approaches to vertebrate ecotoxicity tests in the 21st century: a review of developments over the last 2 decades and current status. *Environmental Toxicology and Chemistry.* **35**, 2637-2646 (2016).
39 Kim, J.-W. et al. Acute toxicity of pharmaceutical and personal care products on freshwater crustacean (Thamnocephalus platyurus) and fish (Oryzias latipes). *Journal of Toxicological Sciences.* **34**, 227-232 (2009).

Figure1

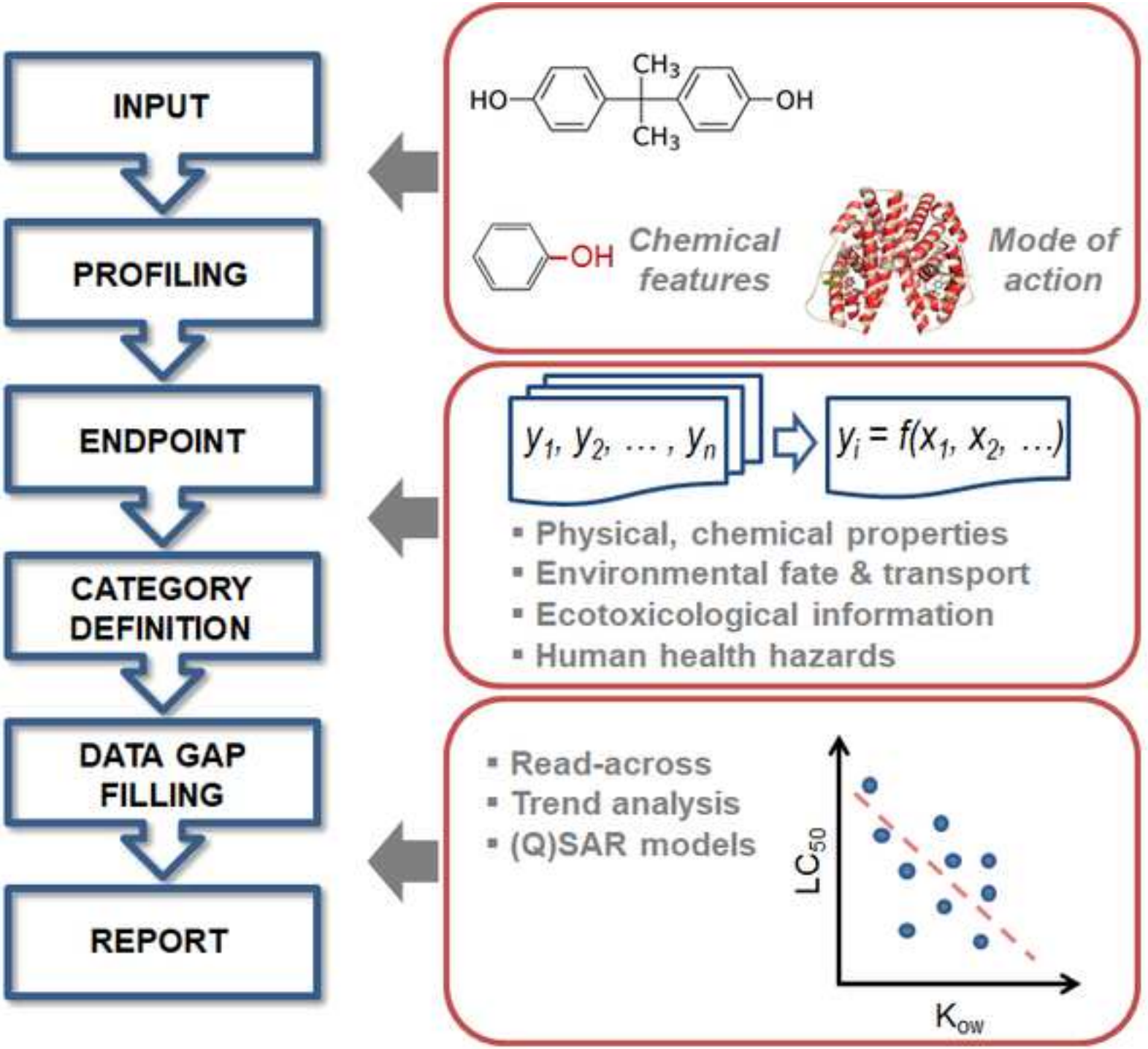


Figure2

Click here to access/download;Figure;Fig2.tif

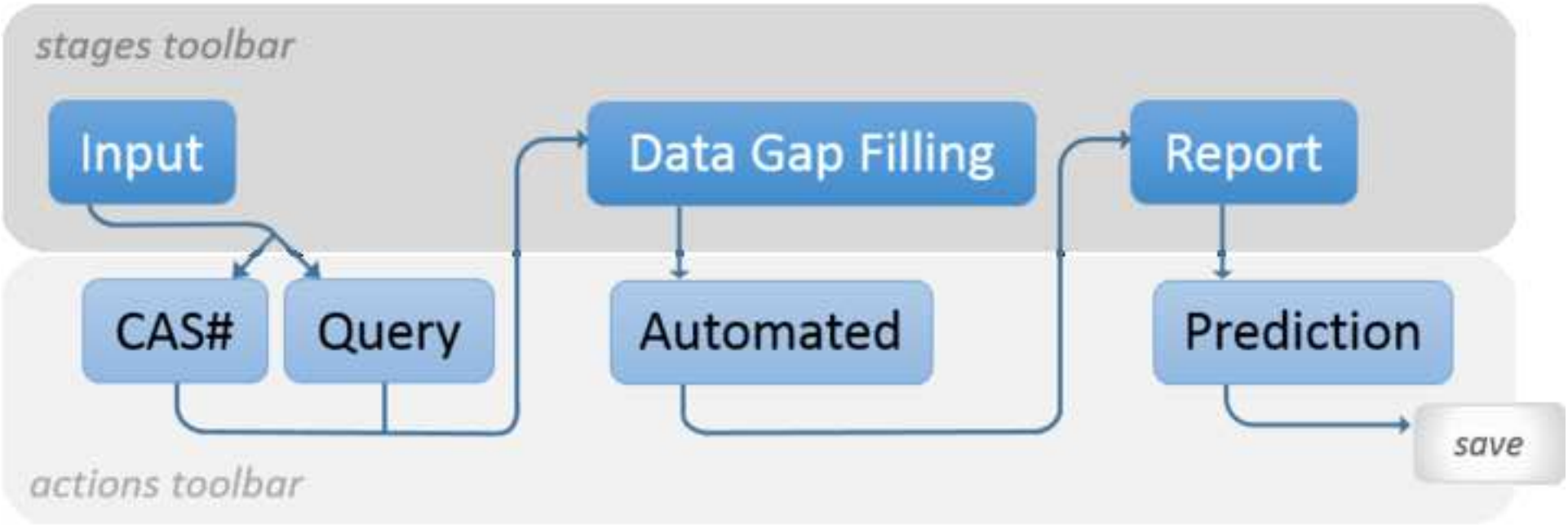


Figure3

Click here to access/download;Figure;Fig3.tif

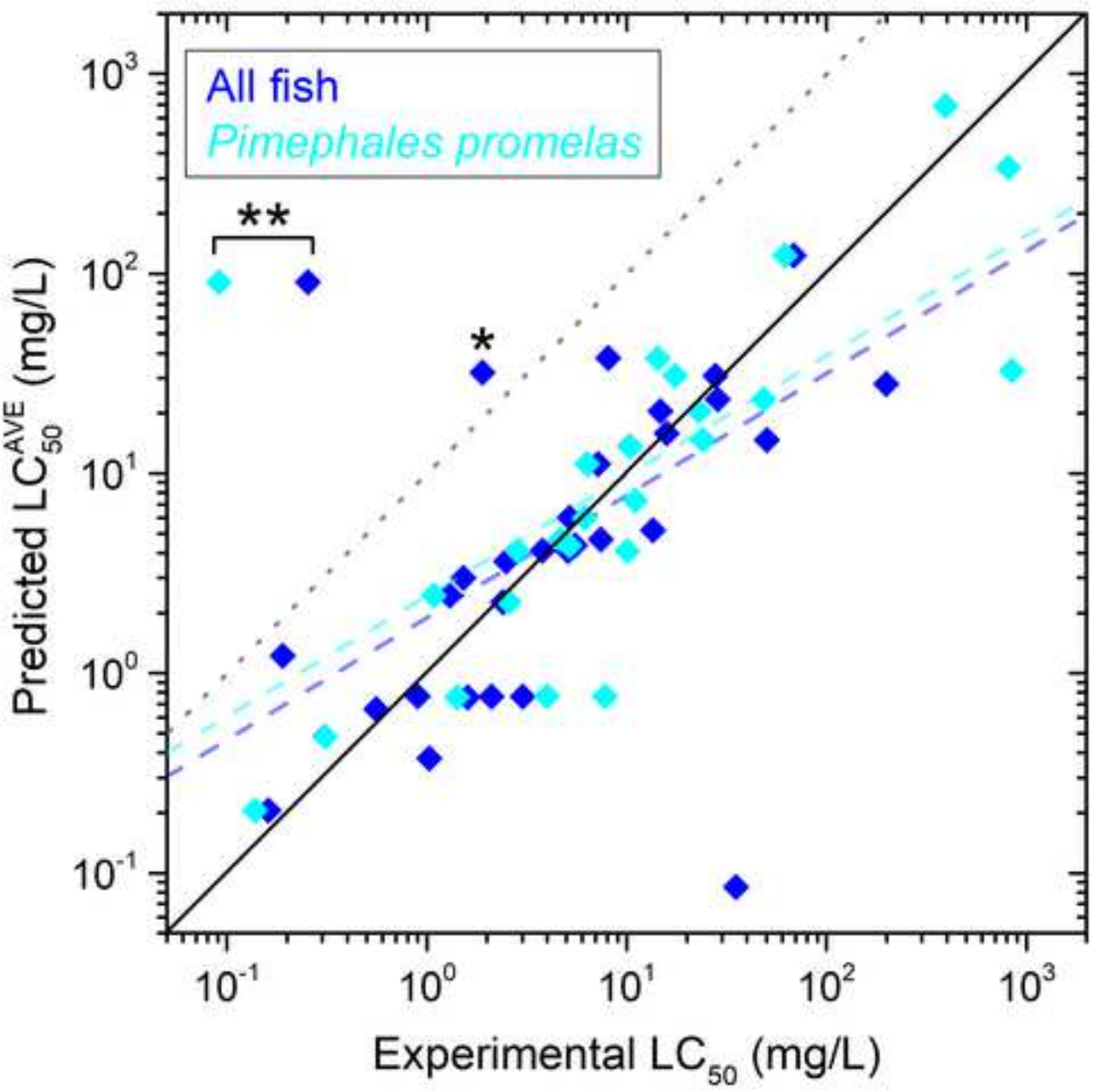


Figure4

Click here to access/download;Figure;Fig4.tif

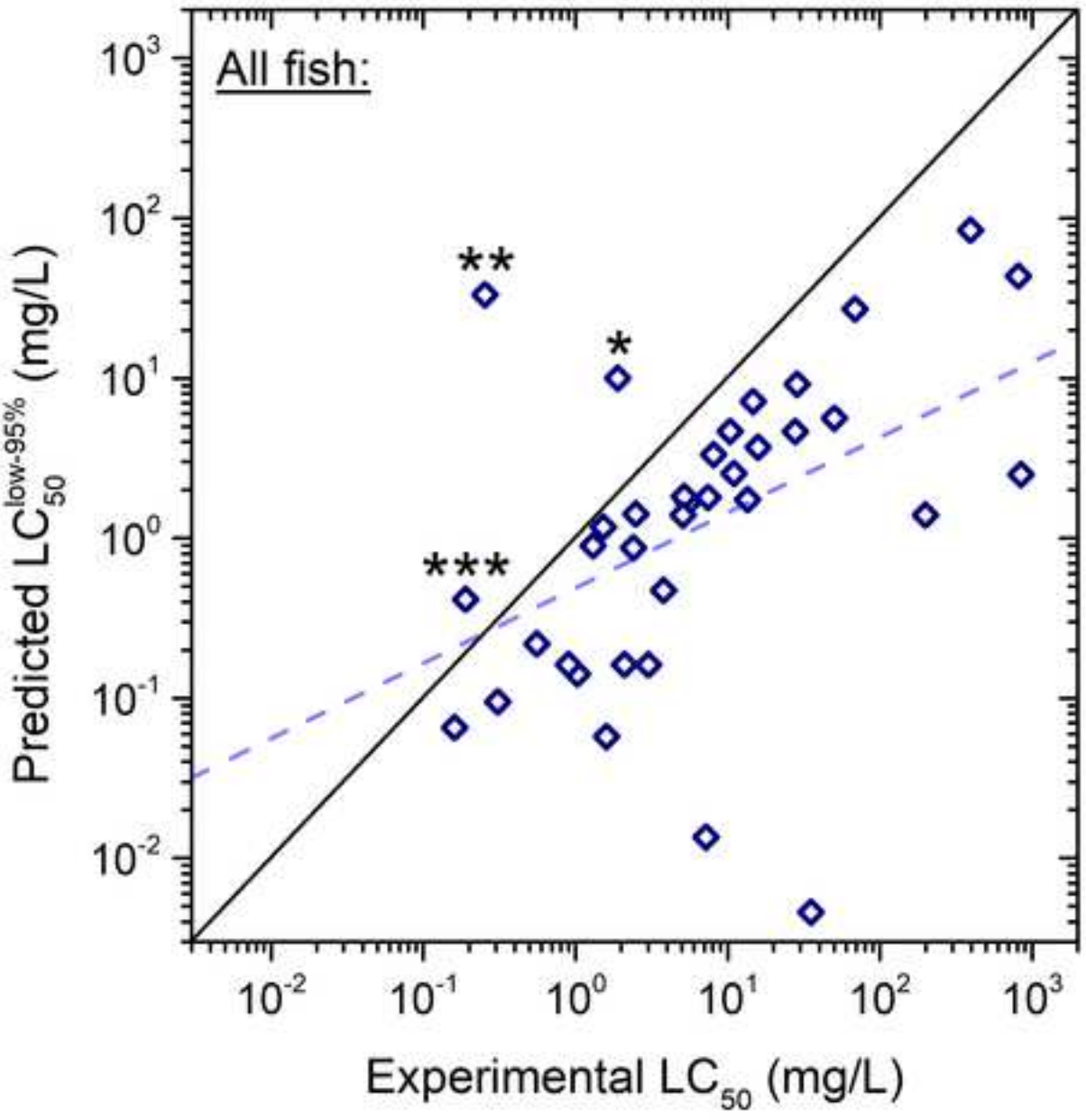
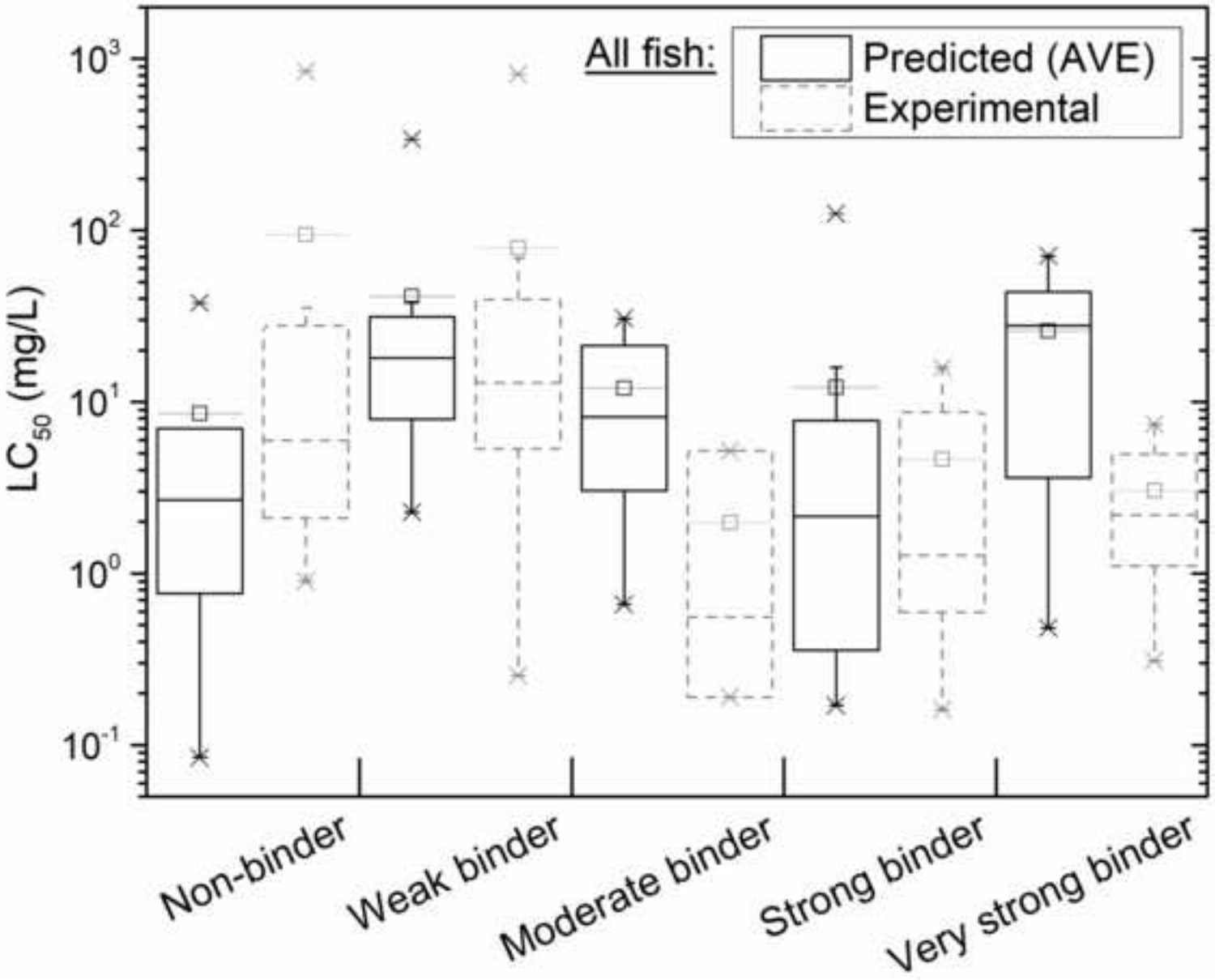


Figure5

Click here to access/download;Figure;Fig5.tif



No.	CAS Registry Number	Substance Name
1	50-28-2	17-β Estradiol
2	57-63-6	17-α Ethinyl-estradiol
3	80-05-7	2,2-bis(4-hydroxyphe-nyl)propane (Bisphenol A)
4	80-46-6	4-tert-Pentylphenol
5	140-66-9	4-tert-Octylphenol
6	446-72-0	Genistein [3',5,7-trihydroxy-4',6-dimethoxyisoflavone]
7	10161-33-8	17β-Trenbolone
8	67747-09-5	Prochloraz (DMI fungicide)
9	84852-15-3	4-Nonylphenol
10	69-72-7	salicylic acid
11	80-09-1	4,4’-dihydroxydiphenyl sulphone (Bisfenol S)
12	84-74-2	phthalic acid, dibutyl ester
13	92-88-6	4,4’-dihydroxybiphenyl
14	94-13-3	4-hydroxybenzoic acid, propyl ester
15	98-54-4	4-tert-butylphenol
16	97-23-4	2,2’-dihydroxy-–5,5’-dichlorodiphenyl-methane
17	97-53-0	eugenol
18	99-76-3	4-hydroxybenzoic acid, methyl ester
19	103-90-2	N-(4-hydroxyphenyl) acetamide
20	106-44-5	p-cresol
21	108-39-4	m-cresol
22	108-45-2	1,3-phenylenediamine
23	108-46-3	1,3-dihydroxybenzene
24	108-91-8	cyclohexylamine
25	119-36-8	salicylic acid, methyl ester
26	120-47-8	4-hydroxybenzoic acid, ethyl ester
27	120-80-9	1,2-dihydroxybenzene
28	123-31-9	1,4-dihydroxybenzene [1,4-benzenediol]
29	131-53-3	2,2’-dihydroxy-4-methoxybenzophenone
30	131-56-6	2,4-dihydroxybenzophenone
31	131-57-7	2-hydroxy-4-methoxybenzophenone
32	599-64-4	4-cumylphenol
33	2855-13-2	1-amino-3-aminomethyl-3,5,5-trimethyl-cyclohexane
34	6864-37-5	3,3’-dimethyl-4,4’-diaminodicyclohexylmethane
35	25013-16-5	tert-butyl-4-hydroxyanisole
36	147315-50-2	2-(4,6-diphenyl-1,3,5-triazin-2-yl)-5-(hexyloxy)phenol
37	88-68-6	2-aminobenzamide
38	611-99-4	4,4’-dihydroxybenzophenone
39	27955-94-8	1,1,1-tris(4-hydroxyphenol)ethane
40	87-18-3	salicylic acid, 4-tert-butylphenyl ester
41	47465-97-4	3,3-bis(3-methyl-4-hydroxyphenyl)2-indolinone
42	99-96-7	p-hydroxybenzoic acid
43	80-07-9	1-Chloro-4-(4-chlorophenyl)sulfonylbenz
44	84-65-1	9,10-Anthraquinone
45	85-44-9	2-benzofuran-1,3-dione

46	92-84-2	10H-Phenothiazine
47	2855-13-2	1-amino-3-aminomethyl-3,5,5-trimethyl-cyclohexane
48	50-27-1	Estriol
49	50-50-0	beta-Estradiol-3-benzoate
50	53-16-7	Estrone
51	92-52-4	Biphenyl
52	92-69-3	p-Phenylphenol
53	96-29-7	2-Butanone oxime
54	121-75-5	Malathon
55	123-07-9	4-Ethylphenol
56	645-56-7	4-n-Propylpehnol
57	1638-22-8	p-Butyl phenol
58	1912-24-9	Atrazine
59	40596-69-8	Methoprene
60	1987-50-4	4-Heptylphenol
61	92-86-4	p,p'-Dibromobiphenyl
62	480-41-1	Naringenin
63	486-66-8	Daidzein
64	491-70-3	Luteolin
65	491-80-5	Biochanin A
66	520-18-3	Kaempferol
67	2051-60-7	2-Chlorobiphenyl (PCB 1)
68	2051-61-8	3-Chlorobiphenyl (PCB 2)
69	2051-62-9	4-Chloro-1,1'-biphenyl
70	2446-69-7	p-n-Hexylphenol [4-hexylphenol]
71	14938-35-3	4-n-Amylphenol
72	17924-92-4	Zearalenone
73	1743-60-8	beta-Estradiol 3-benzoate 17-nbutyrate
74	479-13-0	Coumestrol

SMILES Formula (2D non-stereochemical form)

CC12CCC3C(CCc4cc(O)ccc34)C1CCC2O
CC12CCC3C(CCc4cc(O)ccc34)C1CCC2(O)C#C
CC(C)(c1ccc(O)cc1)c1ccc(O)cc1
CCC(C)(C)c1ccc(O)cc1
CC(C)(C)CC(C)(C)c1ccc(O)cc1
Oc1ccc(cc1)C1=COc2cc(O)cc(O)c2C1=O
CC12C=CC3C(CCC4=CC(=O)CCC=34)C1CCC2O
CCCN(CCOc1c(Cl)cc(Cl)cc1Cl)C(=O)n1ccnc1
CC(C)CCCCCCc1ccc(O)cc1
OC(=O)c1ccccc1O
Oc1ccc(cc1)S(=O)(=O)c1ccc(O)cc1
CCCCOC(=O)c1ccccc1C(=O)OCCCC
Oc1ccc(cc1)-c1ccc(O)cc1
CCCOC(=O)c1ccc(O)cc1
CC(C)(C)c1ccc(O)cc1
Oc1ccc(Cl)cc1Cc1cc(Cl)ccc1O
COc1cc(CC=C)ccc1O
COC(=O)c1ccc(O)cc1
CC(=O)Nc1ccc(O)cc1
Cc1ccc(O)cc1
Cc1cccc(O)c1
Nc1cccc(N)c1
Oc1cccc(O)c1
NC1CCCCC1
COC(=O)c1ccccc1O
CCOC(=O)c1ccc(O)cc1
Oc1ccccc1O
Oc1ccc(O)cc1
COc1ccc(C(=O)c2ccccc2O)c(O)c1
Oc1ccc(c(O)c1)C(=O)c1ccccc1
COc1ccc(C(=O)c2ccccc2)c(O)c1
CC(C)(c1ccccc1)c1ccc(O)cc1
CC1(C)CC(N)CC(C)(CN)C1
CC1CC(CCC1N)CC1CCC(N)C(C)C1
COc1ccc(O)c(c1)C(C)(C)C
CCCCCCOc1ccc(c(O)c1)-c1nc(nc(n1)-c1ccccc1)-c1ccccc1
NC(=O)c1ccccc1N
Oc1ccc(cc1)C(=O)c1ccc(O)cc1
CC(c1ccc(O)cc1)(c1ccc(O)cc1)c1ccc(O)cc1
CC(C)(C)c1ccc(OC(=O)c2ccccc2O)cc1
Cc1cc(ccc1O)C1(C(=O)Nc2ccccc12)c1ccc(O)c(C)c1
OC(=O)c1ccc(O)cc1
Clc1ccc(cc1)S(=O)(=O)c1ccc(Cl)cc1
O=C1c2ccccc2C(=O)c2ccccc12
O=C1OC(=O)c2ccccc12

N1c2ccccc2Sc2ccccc12
CC1(C)CC(N)CC(C)(CN)C1
CC12CCC3C(CCc4cc(O)ccc34)C1CC(O)C2O
CC12CCC3C(CCc4cc(OC(=O)c5ccccc5)ccc34)C1CCC2O
CC12CCC3C(CCc4cc(O)ccc34)C1CCC2=O
c1ccc(cc1)-c1ccccc1
Oc1ccc(cc1)-c1ccccc1
CCC(C)=NO
CCOC(=O)CC(SP(=S)(OC)OC)C(=O)OCC
CCc1ccc(O)cc1
CCCc1ccc(O)cc1
CCCCc1ccc(O)cc1
CCNc1nc(Cl)nc(NC(C)C)n1
COC(C)(C)CCCC(C)CC=CC(C)=CC(=O)OC(C)C
CCCCCCCc1ccc(O)cc1
Brc1ccc(cc1)-c1ccc(Br)cc1
Oc1ccc(cc1)C1CC(=O)c2c(O)cc(O)cc2O1
Oc1ccc(cc1)C1=COc2cc(O)ccc2C1=O
Oc1cc(O)c2C(=O)C=C(Oc2c1)c1ccc(O)c(O)c1
COc1ccc(cc1)C1=COc2cc(O)cc(O)c2C1=O
Oc1ccc(cc1)C1Oc2cc(O)cc(O)c2C(=O)C=1O
Clc1ccccc1-c1ccccc1
Clc1cccc(c1)-c1ccccc1
Clc1ccc(cc1)-c1ccccc1
CCCCCCc1ccc(O)cc1
CCCCCc1ccc(O)cc1
CC1CCCC(=O)CCCC=Cc2cc(O)cc(O)c2C(=O)O1
CC(=O)OC1CCC2C3CCc4cc(O)ccc4C3CCC12C
Oc1ccc2c(OC(=O)c3c-2oc2cc(O)ccc32)c1

Log Kow	AVE predicted 96-h LC50 (mg/L)	LOWER 95% CI predicted 96-h LC50 (mg/L)	Profiler - Estrogen Receptor Binding
4.01	3.62	1.42	Very strong binder, OH group
3.67	3.00	1.18	Strong binder, OH group
3.32	4.68	1.80	Very strong binder, OH group
3.91	2.27	0.87	Weak binder, OH group
5.28	0.38	0.14	Strong binder, OH group
2.84	32.00	10.03	Very strong binder, OH group
2.65	124.72	19.75	Strong binder, OH group
4.1	5.19	1.74	Non binder, without OH or NH2 group
5.92	0.21	0.07	Strong binder, OH group
2.26	24.07	9.31	Weak binder, OH group
1.65	48.67	10.67	Very strong binder, OH group
4.5	0.76	0.06	Non binder, without OH or NH2 group
2.8	12.05	4.20	Moderate binder, OH grooup
3.04	10.32	3.86	Moderate binder, OH grooup
3.31	4.36	1.68	Weak binder, OH group
4.26	0.48	0.10	Very strong binder, OH group
2.27	14.70	5.60	Weak binder, OH group
1.96	38.20	14.01	Weak binder, OH group
0.46	338.97	43.39	Weak binder, OH group
1.94	20.47	7.14	Weak binder, OH group
1.96	23.45	9.17	Weak binder, OH group
-0.33	34.60	0.00	Weak binder, NH2 group
0.8	123.03	27.06	Weak binder, OH group
1.49	28.08	1.40	Weak binder, NH2 group
2.55	16.16	5.68	Weak binder, OH group
2.47	19.93	7.40	Weak binder, OH group
0.88	11.14	0.01	Weak binder, OH group
0.59	90.75	33.19	Weak binder, OH group
3.82	3.97	1.46	Very strong binder, OH group
2.96	12.04	4.73	Strong binder, OH group
3.79	5.96	2.27	Strong binder, OH group
4.12	2.15	0.84	Strong binder, OH group
1.9	30.65	1.53	Moderate binder, NH2 group
4.1	1.07	0.05	Strong binder, NH2 group
3.5	4.85	1.85	Moderate binder, OH grooup
6.24	0.17	0.06	Strong binder, OH group
0.35	694.00	84.30	Weak binder, NH2 group
2.19	37.74	14.67	Very strong binder, OH group
4.38	2.09	0.82	Very strong binder, OH group
5.73	0.24	0.09	Strong binder, OH group
4.48	2.07	0.77	Very strong binder, OH group
1.58	8.54	0.00	Weak binder, OH group
3.9	3.92	0.85	Non binder, without OH or NH2 group
3.39	7.00	3.54	Non binder, without OH or NH2 group
1.6	2.69	0.00	Non binder, without OH or NH2 group

4.15	1.07	0.08	Non binder, without OH or NH2 group
1.9	30.65	1.53	Moderate binder, NH2 group
2.45	21.21	8.29	Very strong binder, OH group
5.47	0.36	0.02	Strong binder, OH group
3.13	7.78	3.06	Strong binder, OH group
4.01	4.10	0.47	Non binder, without OH or NH2 group
3.2	5.99	1.82	Moderate binder, OH grooup
0.63	32.67	2.49	Non binder, non cyclic structure
2.36	37.73	3.33	Non binder, non cyclic structure
2.58	13.63	4.65	Weak binder, OH group
3.2	7.32	2.55	Weak binder, OH group
3.65	4.09	1.39	Weak binder, OH group
2.61	30.87	4.63	Non binder, without OH or NH2 group
5.5	0.08	0.00	Non binder, non cyclic structure
5.01	0.66	0.22	Moderate binder, OH grooup
5.72	0.11	0.02	Non binder, without OH or NH2 group
2.52	27.84	10.87	Very strong binder, OH group
2.55	36.47	11.71	Very strong binder, OH group
2.53	43.75	14.28	Very strong binder, OH group
3.41	15.87	3.70	Strong binder, OH group
1.96	70.98	8.05	Very strong binder, OH group
4.53	0.77	0.16	Non binder, without OH or NH2 group
4.58	0.77	0.16	Non binder, without OH or NH2 group
4.61	0.77	0.16	Non binder, without OH or NH2 group
4.52	1.22	0.42	Moderate binder, OH grooup
4.06	2.44	0.89	Weak binder, OH group
3.58	7.22	2.66	Strong binder, OH group
4.95	0.91	0.35	Strong binder, OH group
1.57	52.16	11.44	Very strong binder, OH group

Name of Material/ Equipment	Company	Catalog Number	Comments/Description
Adobe Acrobat Reader DC	Adobe Systems Software Ireland Limited	NA	Required to view prediction and category report
Computer	System: Microsoft Corporation	NA	Recommended system properties: (i) system type: 64 bit, Microsoft Windows 7 or newer, (ii) processor: I5 at 2.4 GHz or faster processor or equivalent AMD CPU, (iii) Installed memory (RAM): 6 GB of RAM, (iv) Hard Disk Drive (HDD): 20 GB free hard drive space
Microsoft Editor	Microsoft Corporation	NA	Required to upload a substance list of CAS numbers (batch mode) to the OECD QSAR Toolbox as .txt file (text file)
Microsoft Excel 2016	Microsoft Corporation	NA	Required to export data from OECD QSAR Toolbox as .cvs, .xls or .xlsx files
OECD QSAR Toolbox version 4.0 or newer	Organisation for Economic Co-operation and Development	NA	Required to run OECD QSAR Toolbox Automated Workflows; free download: https://qsartoolbox.org/download/
OriginPro 9	OriginLab Corporation	NA	Optional program for data analysis; similar tools possible

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: In silico modeling method for computational aquatic toxicology ...

Author(s): Marie-Léonie Bohlen, Hynn Pyo Jeon, Young Jun Kim, Boreckyoung Sung

Item 1: The Author elects to have the Materials be made available (as described at http://www.jove.com/publish) via:

☐ Standard Access ☒ Open Access

Item 2: Please select one of the following items:

☒ The Author is **NOT** a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: "**Agreement**" means this Article and Video License Agreement; "**Article**" means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; "**Author**" means the author who is a signatory to this Agreement; "**Collective Work**" means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; "**CRC License**" means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; "**Derivative Work**" means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; "**Institution**" means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; "**JoVE**" means MyJove Corporation, a Massachusetts corporation and the publisher of The Journal of Visualized Experiments; "**Materials**" means the Article and / or the Video; "**Parties**" means the Author and JoVE; "**Video**" means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and(c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

4. **Retention of Rights in Article.** Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. **Grant of Rights in Video – Standard Access.** This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. **Grant of Rights in Video – Open Access.** This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this **Section 6** is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. **Government Employees.** If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. **Protection of the Work.** The Author(s) authorize JoVE to take steps in the Author(s) name and on their behalf if JoVE believes some third party could be infringing or might infringe the copyright of either the Author's Article and/or Video.

9. **Likeness, Privacy, Personality.** The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

10. **Author Warranties.** The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

11. **JoVE Discretion.** If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

12. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

13. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication of the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

14. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement is required per submission.

CORRESPONDING AUTHOR

Name:	Baeckkyoung Sung
Department:	Environmental Safety
Institution:	KIST Europe
Title:	Senior Researcher
Signature:	성백경
Date:	28/03/2019

Please submit a **signed** and **dated** copy of this license by one of the following three methods:

1. Upload an electronic version on the JoVE submission site
2. Fax the document to +1.866.381.2236
3. Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02140

Response to the Reviewers

We thank the reviewers for their invaluable comments and suggestions and hope that our response will adequately address the perceived ambiguities in our manuscript. We believe our responses to the reviewers' questions/comments have improved the manuscript significantly. All changes in the revised manuscript have been marked by using "track changes" in MS Word.

Editorial comments:

General:

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.

RESPONSE: The spelling and grammar have been checked once more.

Protocol:

1. 'TIP', 'INFO', etc. are not part of JoVE format. Please rewrite as numbered protocol steps (in the imperative) or 'Notes'.

RESPONSE: All the TIPs and INFOs have been corrected accordingly.

2. Please add more details to your protocol steps. Please ensure you answer the "how" question, i.e., how is the step performed? Alternatively, add references to published material specifying how to perform the protocol action. If revisions cause a step to have more than 2-3 actions and 4 sentences per step, please split into separate steps or substeps.

RESPONSE: We have revise the manuscript to describe the protocol more in detail.

References:

1. Please do not abbreviate journal titles.

RESPONSE: The "QSAR Toolbox" is an official name of the software which we report in this paper. However, the "OECD" can be omitted in the title. We would keep the name, "QSAR Toolbox", like many other papers where the same abbreviated software name has been used in the title.

Table of Materials:

1. Please ensure the Table of Materials has information on all materials and equipment used, especially those mentioned in the Protocol.

RESPONSE: It has been checked again and described correspondingly.

Reviewers' comments:

Reviewer #1:

Major Concerns:

There is no novelty in showing a "reasonable" correlation between the toxicity of these compounds and log Kow.

RESPONSE: The authors agree with this point. The graphs showing the relationships between log Kow and LC50 & ER binding have been moved to Supplementary Information and the Discussion part has been revised correspondingly (the last 3 paragraphs in Discussion). We would emphasize that the aim of our paper is to describe the protocol on how to use the QSAR Toolbox software to quantitatively analyze the toxic effects of endocrine disrupting chemicals, and the demonstrated process to produce the data relating to log Kow is valid. However, since the critique of the reviewer is correct, we stressed the limitation of this approach in the Discussion (the last paragraph). We would mention that the presented methodology here, for analyzing toxicity of EDs by using OECD QSAR Toolbox, has not yet been reported as a protocol paper as far as we know.

The discussion of outliers (e.g. lines 366-376) does not take account of known mechanisms of action - go and look at the original publication of these data from the US EPA e.g. Russom, C.L., S.P. Bradbury, S.J. Broderius, D.E. Hammermeister and R.A. Drummond. 1997. Predicting modes of action from chemical structure: Acute toxicity in the fathead minnow (Pimephales promelas). Environmental Toxicology and Chemistry 16(5): 948-967 and the references therein.

RESPONSE: The authors agree with this point and reinforced the discussion on the mechanism of action of the outliers (the 4th & 5th paragraphs in Discussion), citing the above paper. Thank you for suggesting a valuable reference.

Reviewer #2:

Manuscript Summary:

Review of: "In Silico Modeling Method for Computational Aquatic Toxicology of Endocrine Disruptors: A software-based approach for using the OECD QSAR Toolbox"

The authors report on the application of the QSAR toolbox to compare ecotoxicological LC50 values for fish (concentration for 50% of lethality) with logKow values. The conclusion of the paper, not surprisingly, is that there is little relationship between the two but the paper is intending to serve the role of demonstrating to the reader how to use the toolbox to perform such analyses. For clarity, the reviewer did not run through the described protocol in the paper so cannot confirm that it runs as described. However, the general review or "vignette" is an appropriate description of HOW to use the

dashboard for this purpose, whether the study itself is of value and whether or not the results are surprising or not.

RESPONSE: The authors do agree with the opinion of the reviewer. The graphs showing the relationships between log Kow and LC50 & ER binding have been moved to Supplementary Information and the Discussion part has been revised correspondingly (the last 3 paragraphs in Discussion), with stressing the limitation of this approach in the Discussion (the last paragraph). We adopted the most of the reviewer´s view-point in our revision, which is highly appropriate and correct, and would appreciate the reviewer for the valuable comments.

I believe the scientific basis for the study of fish LC50 values vs logKow for endocrine disruption is, at best, a weak driver for the publication as there is enough literature reported examining such relationships. However, the application of the QSAR Toolbox to investigate the correlation is a valid application and a publication on the vignette may well be of value but the reviewer questions whether such an approach may already be captured in the QSAR Toolbox manual?

RESPONSE: The authors fully agree with this point. In accordance to this viewpoint, we have removed the plots on log Kow from the main text, and more stressed on the other data, i.e., (1) the predicted vs. experimental LC50 and (2) the LC50 vs. ER binding affinity. The unrealistic correlation between log Kow and LC50 & ER binding has been clarified in the last paragraphs of Discussion, in accordance to the reviewer´s comments. In spite of such limitation, as the reviewer points out, we believe that our protocol and results are valid to be used as a vignette for one of the applications of QSAR Toolbox. It should be noted that, different from our protocol, the QSAR Toolbox manual only focuses on overall and functional presentation on the workflow. In contrast, our protocol highlights critical steps during the operation of algorithms (e.g., potential pitfalls and core advice in profiling and data gap filling) and where to find the relevant information on substance characteristics in a user-friendly manner, which allows improved reproducibility in data generation. The authors do thank the reviewer for the advice and guidance.

Major Concerns:

Line 98-100: The paper does NOT demonstrate that the QSAR Toolbox can predict the "reproductive toxicities" of aquatic vertebrates! LC50 is looking at lethality/mortality - not reproductive tox. It tries to identify a correlation between Kow and LC50 so this extrapolation is extreme.

RESPONSE: The authors agree with this comment and it is reflected in the revised version, as explained in the above responses.

Minor Concerns:

Line 122: I do not rate them as automated workflows - they are stage by stage human driven.

RESPONSE: The entire stages indeed cannot be said to be automated; however, in our protocol, the computation for the prediction is fully automated and takes place in stage of “Data gap filling”. In fact, the term of “automated workflow” has been defined by the developers of QSAR Toolbox software in this

context. The other stages are also necessary to be run on the “automated workflow”, but are actually human driven, as the reviewer points out. At this point, our protocol is found to be useful, because the “automated workflow” needs some manual preparations in order to be executed.

Line 131: So the workflow fails if there is no structure available? There is no way to enter a structure only use lookup?

RESPONSE: Yes, if no SMILES is available, the QSAR Toolbox cannot do any computation. It is possible to draw a structure, and on the basis of the drawn structure a SMILES will be generated. The Toolbox will search for the SMILES in the integrated databases and inventories, and the correct record can be chosen. If the structure cannot be retrieved in any database and inventory, the prediction can still be run based on the SMILES.

Line 138: Why are multiple structures returned for a single CASRN? The mapping should be one to one.

RESPONSE: CAS numbers are not unique identifiers. Under REACH for example, slightly different substances, e.g. with regard to their purity, are collectively assigned to a single CAS number. As the QSAR Toolbox is connected to ECHA CHEM, multiple records can be retrieved for a single CAS number. SMILES codes on the other hand a unique identifier.

Line 375: Why is applying 3D QSAR modeling deemed to be a necessity? To describe outliers? There can be a myriad of reasons for outliers that will not be solved by modeling 3D QSAR

RESPONSE: The corresponding part, mentioning 3D QSAR, has been removed in the revised manuscript.

Line 381: I do not think of SMILES as 2D descriptors of structures - they are linear notation representations of the structure whereas descriptors would be those represented by Padel, DRAGON etc.

RESPONSE: We agree. The “SMILES” has been deleted from the sentence.

Line 387-392: This is the real reason why the approach is failing - logKow is not a good direct predictor of aquatic LC50 and this is well proven.

RESPONSE: We agree. The revision has been performed according to the reviewer´s comment, as has been explained in the first two responses.

The list of SMILES formulae should be noted to be for 2D non-stereochemical containing forms.

RESPONSE: SMILES can be searched for as 2D non-stereochemical or 3D stereochemical containing forms in the QSAR Toolbox. The basis of computation is the 2D non-stereochemical form. The corresponding parts and Table 1 header have been corrected based on the reviewer’s comment. Thank you for this comment.

The table of experimental versus predicted logKow values should not list values to NINE decimal places. 2 would suffice.

RESPONSE: The values have been corrected to exhibit two decimals.

Some minor comments

Line 53:

There are some minor edits in the text required

Line 46/47: The dependence of the estrogen receptor binding affinity on the relationship between Kow and LC50 is also discussed.

Line 54: of A given chemicals database

Line 56: QSAR tools CAN PRODUCE predictive toxicity profiles

Line 88: "Since the release of THE first version"

Line 105: straightforward

RESPONSE: All the minor comments have been reflected in the revision. Thank you for the details.

Reviewer #3:

Manuscript Summary:

This article describes a method to assess the toxicity to fish of endocrine disrupting chemicals using the tools implemented in OECD QSAR Toolbox.

Major Concerns:

Summary of major concerns: Even if the general methodology presented here is correct, the proposition of authors to correlate acute toxicity to fish (lethality) to ER binding seems inappropriate. And they indeed don't find a significant correlation. Studying the link between ER binding and chronic fish toxicity would be relevant, and a correlation might be found (which is only suggested in the last sentence of the manuscript). I am also concerned by the substances included in the dataset, as some are known EDCs because they bind to ER, but some others are EDCs because they bind to AR. Some other are probably not EDCs. A thorough validation of this dataset could be useful.

RESPONSE: The authors agree with this point. The graphs showing the relationships between log Kow and LC50 & ER binding have been moved to Supplementary Information and the Discussion has been revised correspondingly (the last 3 paragraphs in Discussion). The unrealistic correlation between log Kow and LC50 & ER binding has been clarified in the last paragraphs of Discussion, in accordance to the reviewer´s comments. By using QSAR Toolbox, we have extensively tried to find a way to generate the data on chronic toxicity of fish with the list of EDs, but the current version of the Toolbox could produce few long-term predictions, and this point has been mentioned together with the future directions for improvement. Furthermore, the QSAR Toolbox does not provide an analysis option for AR binding. Our list of EDs has been adopted from those of the three previous papers (Nendza et al., 2019; Roncaglioni et al., 2008; Sosnová et al., 2016; note, references are provided in table caption of Table 1) where the

same topic has been analyzed by using different methodologies. However, our list of EDs has been partially modified according to the reviewer´s comment, as explained below.

line 250: Some of the compounds in table 1 are certainly not EDCs (e.g. the cresols, or the phenol), some of them are EDCs, but not related to ER (e.g. flutamide, which is metabolised into hydroxyflutamide binding to AR, but not ER). Therefore, you have a non-homogenous dataset, which may lead to erroneous analysis of the data and conclusions.

RESPONSE: The authors agree with this point. The three chemicals (cresol, phenol, and flutamide) have been removed from our original list of EDs, and all the graphs have been re-plotted and fitted in accordance to this change.

lines 283-285: The impact of ER binding on the acute fish toxicity is not very significant, as shown in figure 7. To see the impact of ER binding on toxicity, the authors should rather compare the chronic toxicity to fish to ER binding. A strong ER binding will impact the reproduction and the development of fish, but the lethality of parents in an acute test is not due to ER binding, but rather to narcosis in most cases, or to RedOx cycling in the case of 1,4-benzenediol (hydroquinone) for example.

RESPONSE: The authors do agree with the necessity of the analysis of chronic toxicity for EDs, which is not yet available as “automated workflow” in the current version of QSAR Toolbox. In fact, there is no standard and straightforward workflow available for aquatic long-term toxicity in the current version. Prediction can be performed but fully manual. In addition, prediction of aquatic long-term toxicity is limited and often cannot be produced based on too little available data.

Furthermore, we encountered in our preliminary study that using effect concentrations from long-term aquatic toxicity studies showed far higher effect concentrations than effect concentrations related to purely endocrine endpoints, such as the effects on the number of offspring compared to change in hormone levels. Therefore, predicted long-term toxicity data, which we generated in our preliminary study in the QSAR Toolbox, was not comparable to observe no effect concentrations (NOECs) of purely endocrine endpoints. The reason behind is that endpoints such as number of offspring, which is commonly measured in long-term toxicity studies and therefore available data in the QSAR toolbox, is often less sensitive compared to change in hormone levels. Therefore, comparing predicted results based on less sensitive endpoints were not well comparable to the previously published effect concentrations of the respective endocrine disrupter.

In the revised manuscript, this limitation has been mentioned in the last paragraphs of Discussion, reflecting the reviewer’s comment. Thank you for valuable advice.

Minor Concerns:

lines 56-58: this definition of QSARs seems over-complicated, difficult to understand to me, while the principle of a QSAR is not that complicated. Possible reformulation: QSAR models link descriptor values (e.g. physicochemical or molecular parameters) to a toxicity endpoint values (e.g. LC50 = lethal concentration for 50% of test organisms). Then, the QSAR model can be used to predict this endpoint value for a new substance based on the values of the descriptors for this substance.

RESPONSE: The authors fully agree with this and the corresponding part has been revised according to the reviewer's comment.

line 60: To be an operational and robust QSAR, it does not necessarily have a high-volume dataset. Better have 20 points with high-quality experimental studies than 1000 points with low-quality studies.

RESPONSE: The authors fully agree with this and the corresponding part has been revised according to the reviewer's comment.

line 142: For batch mode input file: Don't you need to save the text file as .smi instead of .txt or the tool can not regognise it? I suggest also to write something like: "Refer to OECD QSAR Toolbox help file for guidelines related to writing of a batch input file."

RESPONSE: No, the tool actually needs a list of CAS numbers saved as text file. Unfortunately, there is no guidance provided on how to prepare the text file in the manual. In the Supplementary Information, we have added an example text file (Supplementary Fig. S3).

line 167: Please add a reference to the modified Verhaar scheme, i.e. Verhaar et al., 1992, Enoch et al., 2008.

RESPONSE: The references have been added. Thank you for the suggestion.

lines 167-168: "provides a first estimate of the acute aquatic toxicity". Rather put "a first estimate of the acute aquatic toxicity mechanism". Verhaar scheme doesn't give an estimate of the toxicity (i.e. quantitatively), but just by which mechanism toxicity occurs. A baseline narcotic can be very toxic (if very hydrophobic) or very harmless (if very hydrophillic).

RESPONSE: The corresponding part has been corrected based on the reviewer's comment. Thank you for valuable comments.

lines 373-376: The mechanism of toxicity of hydroquinone is well known and docking or 3D QSAR modelling doesn't appear "a necessity". The main reason for its acute toxicity is RedOx cycling: it is oxidized to benzoquinone and reduced back to semi-quinone or hydroquinone again and again, depleting cofactors and generating ROS. See (Bolton et al, 2000) "Role of Quinones in Toxicology".

RESPONSE: The corresponding part in Discussion has been corrected based on the reviewer's comment. Thank you for valuable advice and reference.

line 385: The receptor binding is also due to hydrogen-bonding and pi-stacking. And the position of each chemical group on the molecule is also very important. If you move the OH groups of estradiol at other positions on the molecule, you will lose the receptor affinity, even if the hydrrophobicity and number of hydrogen-bond acceptors/donors remain the same.

RESPONSE: The corresponding part in Discussion has been corrected based on the reviewer's comment. We appreciate the reviewer for the valuable advice.

lines 389-391: no link because, as explained earlier, this adverse outcome (acute lethality) is probably not due to this molecular initiating event (ER binding). But if you take developmental toxicity in a chronic fish study, then you may see the link between ER binding and toxic outcome.

RESPONSE: The corresponding part has been corrected based on the reviewer's comment.

line 3 of Table 1: please remove the hyphen: "phenyl" instead of "phe-nyl".

RESPONSE: Done. Thank you for the detail.

Supplementary Information

In Silico Modeling Method for Computational Aquatic Toxicology of Endocrine Disruptors: A Software-Based Approach Using QSAR Toolbox

Marie-Léonie Bohlen, Hyun Pyo Jeon, Young Jun Kim, and Baeckkyoung Sung*

KIST Europe Forschungsgesellschaft mbH, 66123 Saarbrücken, Germany

*Correspondence: sung@kist-europe.de

Fig. S1 and the following Eqs. 1 and 2 show the statistical relationship between the descriptor (K_{OW}) and endpoint (LC_{50}) of the EDs listed in Table 1. In the log-log plot, the linear regression for the predicted and experimental values showed a comparable slope ($\log_{10}LC_{50}$ / $\log_{10}K_{OW}$ = -0.509 and -0.414, respectively) with a relatively high coefficient of determination for the prediction (adjusted r^2 = 0.805 and 0.437 for prediction and observation, respectively).

Computational prediction: $$log_{10}LC_{50} = -0.509 \cdot log_{10}K_{ow} + 2.43 \quad \text{(Eq. 1)}$$

Experimental observation: $$log_{10}LC_{50} = -0.414 \cdot log_{10}K_{ow} + 2.12 \quad \text{(Eq. 2)}$$

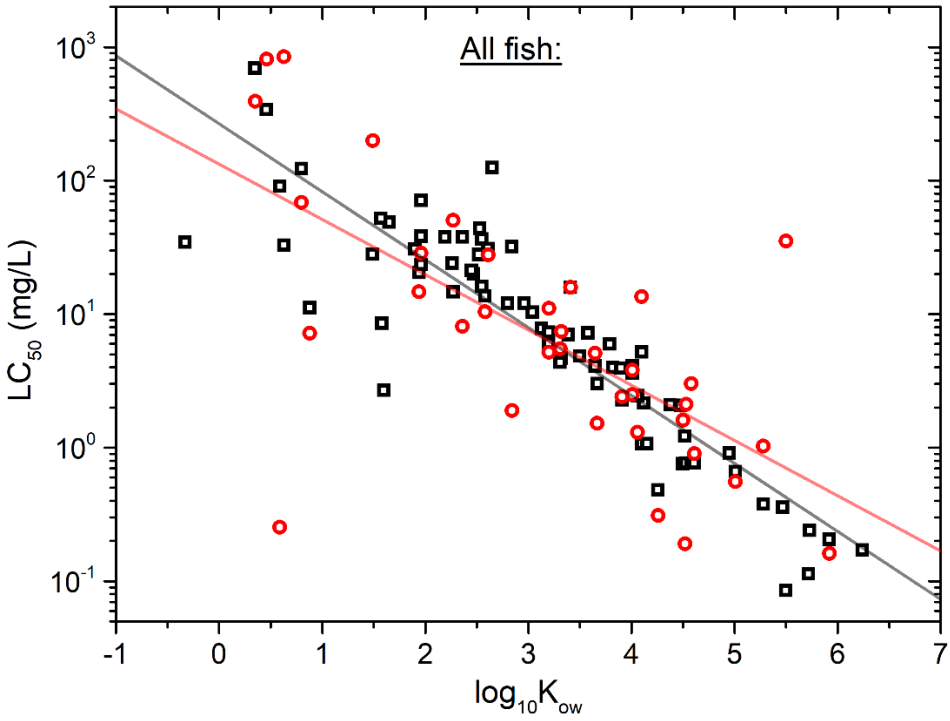


Figure S1: Relationships between 96-h LC_{50} and $\log_{10}K_{OW}$ values of EDs listed in Table 1. For all fish species, the values for predicted (black empty squares, n = 74) and experimental (red empty circles, n = 36) are plotted together in a log scale as a function of $\log_{10}K_{OW}$. For the predicted LC_{50}, the average values are displayed. The gray (Eq. 1) and light-red (Eq. 2) lines represent the linear regression for each case (predicted and experimental, respectively).

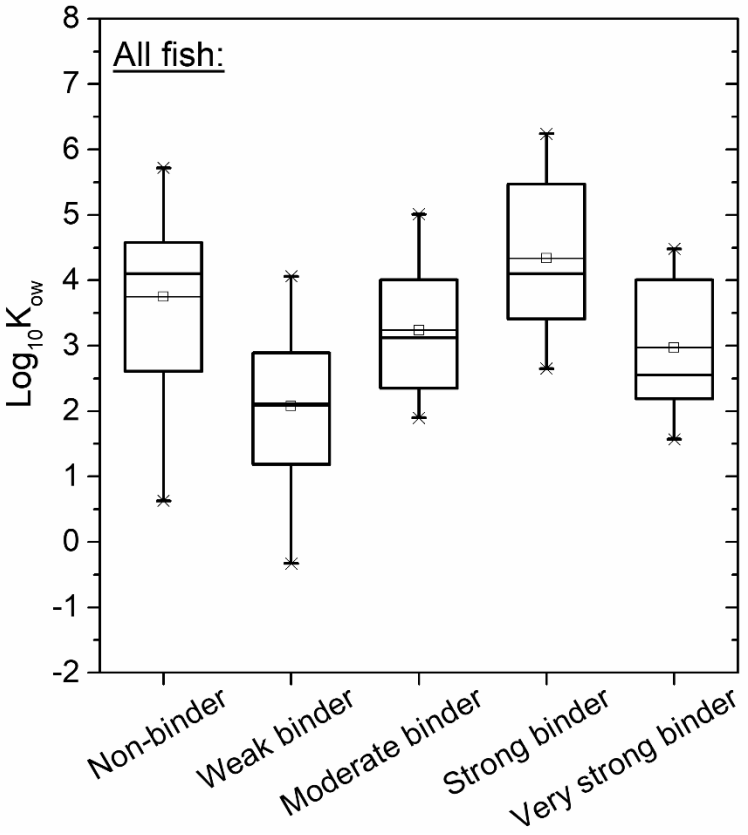


Figure S2: Distributions of the $\log_{10}K_{OW}$ depending on the estrogen receptor (ER) binding affinity of EDs in Table 1 for all fish (n = 8-20 for each category). A box plot represents: (a) mean (small square with a horizontal bar), (b) 1st and 3rd quartiles (lower and upper ends of the box, respectively), (c) median (horizontal segment inside the box), (d) 5th and 95th percentile (lower and upper error bars, respectively), (e) 1th and 99th percentile (lower and upper ×, respectively), and (f) minimum and maximum (lower and upper −, respectively).

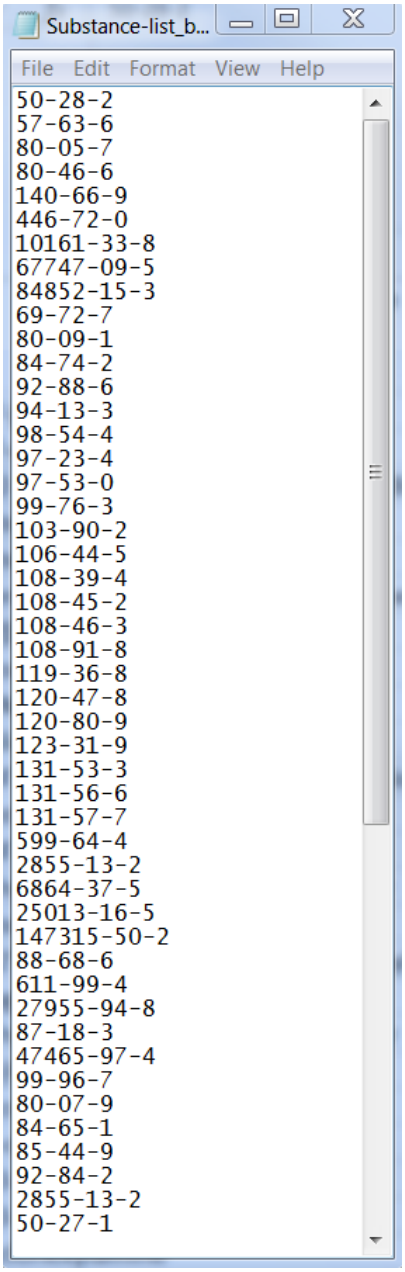


Figure S3: Example of a substance list of CAS numbers for batch mode processing. Each CAS number is listed in a single row. The file was prepared in a text editor and must be saved as text file (.txt).

TITLE:

In Silico Modeling Method for Computational Aquatic Toxicology of Endocrine Disruptors: A Software-Based Approach Using ~~OECD~~ QSAR Toolbox

AUTHORS AND AFFILIATIONS:

Marie-Léonie Bohlen[1], Hyun Pyo Jeon[1], Young Jun Kim[1], and Baeckkyoung Sung[1]

[1]KIST Europe Forschungsgesellschaft mbH, 66123 Saarbrücken, Germany

Corresponding Author:
Baeckkyoung Sung
sung@kist-europe.de
Tel: +49 (0)681-9382-379

Email Addresses of Co-authors:
Marie-Léonie Bohlen (ml.bohlen@kist-europe.de)
Hyun Pyo Jeon (hpjeon@kist-europe.de)
Young Jun Kim (youngjunkim@kist-europe.de)

Field Code Changed

Field Code Changed

Field Code Changed

KEYWORDS:

OECD QSAR Toolbox automated workflow, quantitative structure-activity relationship (QSAR), endocrine disrupting chemical, aquatic vertebrate, acute toxicity, computational ecotoxicology

SUMMARY:

Quantitative structure-activity relationship (QSAR) modeling is a representative bioinformatics-assisted method in toxicological screening. We demonstrate a means of computationally assessing the risks of endocrine disruptors (EDs) in aquatic environments. ~~-~~Utilizing the OECD QSAR Toolbox, we show a protocol to implement *in silico* assay for analyzing toxicity of EDs on fish.

ABSTRACT:

Computational analyses of toxicological processes have enabled high-throughput screening of chemical substances and prediction of their endpoints in biological systems. In particular, quantitative structure-activity relationship (QSAR) models have been increasingly applied to assess the environmental effects of a plethora of toxic materials. In recent years, some of the most highlighted types of toxicants are the endocrine disruptors (EDs) (i.e., chemicals that can interfere with any type of hormone-related metabolism). Because EDs may significantly affect animal development and reproduction, rapidly predicting the adverse effects of EDs using *in silico* techniques is required. In this study, we present an *in silico* method to generate prediction data on the effects of representative EDs on aquatic vertebrates, particularly fish species. We show an example protocol by utilizing the automated workflow of the QSAR Toolbox software developed by the Organisation for Economic Co-operation and Development to enable acute ecotoxicity predictions of EDs. As a result of this protocol, (1) t~~T~~he numerical correlations between the concentration for 50% of lethality (LC_{50}) and the octanol-water partition coefficient

(K_{ow}) are calculated, (2)~~and~~ the output performances are presented in which the LC_{50} values determined in experiments are compared with those generated by computations~~.~~, and (3) t~~T~~he dependence~~s~~ of ~~K_{ow} and LC_{50} on~~ the estrogen receptor binding affinity on the relationship between K_{ow} and LC_{50} ~~are~~ is ~~also analyzed~~ discussed.

INTRODUCTION:

New developments in informatics and computational technology have empowered all areas of biological sciences with quantitative methodologies that offer high precision and reliability[1]. In particular, algorithms used in molecular taxonomy and property classification have resulted in quantitative structure-activity relationship (QSAR) models[2], which automatically correlate the chemical structures and biological activities of a given chemicals database and implement rapid *in silico* screening of a wide range of chemical substrates according to their medicinal or toxicological actions[3]. QSAR tools ~~calculate~~ can produce predictive toxicity profiles as a function of feature vectors of molecular descriptors (i.e., physicochemical parameters) of chemicals of interest to create categorical endpoints numerically[4]. Usually, each quantitative endpoint is displayed as a 2D scatter plot versus changes in descriptor values. A QSAR model is then generated from this by using (multiple) linear regression analyses. Once a ~~high-volume~~ dataset has been fully exploited to construct a QSAR model (called the training set), then the model is statistically validated by predicting the endpoints of a group of chemicals not included in the training set (called the test set). The model can then be used finally to predict the biological activities of untested compounds[3].

Among many harmful chemicals, endocrine disruptors (EDs) have been highlighted as a group of toxicants that may interfere in numerous hormone-related metabolisms in mammals, amphibians, and fish[5,6]. EDs are known to induce a variety of adverse effects, such as cancers and malformations, by blocking or altering normal hormonal pathways or by activating abnormal hormone synthesis/degradation signals. As a consequence, these hormone-mimicking chemicals can perturb endocrine systems such that biological development and reproduction of wildlife animal populations are hampered. In particular, the ecotoxicological effects of EDs have been extensively investigated in aquatic vertebrates, which have nearly identical hormone receptor structures to those of mammals, including humans. Because all hormonal actions occur at low doses *in vivo*, predicting the potential toxicities of ED candidates using rapid *in silico* screening is critical to public and environmental health.

QSAR modeling studies on the toxicology of EDs have been conducted by utilizing both 2D and 3D descriptors (known as 2D and 3D QSAR, respectively), which reveal the ED ligand binding affinities of estrogen, androgen, and progesterone receptors[7]. Despite the high-precision advantages of 3D QSAR, in which conformational and electrostatic interactions are considered, 2D QSAR retains its own robustness in direct mathematical algorithms, rapid calculations, and extremely low computational loads. In addition, 2D-QSAR models are flexible for use in a wide range of applications while achieving relatively accurate prediction performance.

The Organisation for Economic Co-operation and Development (OECD) QSAR Toolbox is currently one of the most utilized computer software tools, providing freely available and pre-built QSAR

models[8,9]. Its profiler uses 2D descriptor databases. Since the release of ~~its~~ the first version in 2008, the software has been applied in the fields of chemical and bio-industry, public health, and environmental safety for full or partial analysis of the potential risks of natural and synthetic compounds, with special interests in carcinogenesis[10-12], mutagenicity[13-15], and developmental toxicity[16]. The application to aquatic toxicology has also been demonstrated with a focus on bioaccumulation and biotransformation[17].

The QSAR Toolbox has been proven useful in predicting the short-term toxicity of broad range of chemicals[17] as well as the estrogen receptor (ER) binding affinities of EDs[18]. However, the acute ecotoxicities of EDs in aquatic vertebrates has not been analyzed using the QSAR Toolbox. In this study, we present a typical and facile protocol to perform QSAR modeling on the acute adverse effects of EDs with a focus on fish species. The study shows that the QSAR Toolbox is a highly-accessible ~~and powerful piece of~~ software for calculating and predicting the ~~reproductive toxicities~~ lethality/mortality of aquatic vertebrates for some representative EDs. The statistical treatment methods for the derived *in silico* datasets are presented.

PROTOCOL:

Fig. 1 shows the overall scheme for the general operation of the QSAR Toolbox. The flow shown in Fig. 2 provides straight-forward instructions on how to operate the *in silico* assay to predict acute ecotoxicity of target substances such as endocrine disrupting chemicals.

1. Equipment

Software: OECD QSAR Toolbox 4.0 or newer (free download from https://qsartoolbox.org/download/) and Microsoft Excel 2016, OriginPro 9, or other similar data analytical tools.

Computer: The recommended system properties for the OECD QSAR Toolbox are: (i) system type: 64 bit, Windows 7 or newer; (ii) processor: I5 at 2.4 GHz, or a faster processor or equivalent AMD CPU; (iii) installed memory (RAM): 6 GB; (iv) Hard Disk Drive (HDD): 20 GB of free hard drive space (OECD QSAR Toolbox 4.3 Release Notes: https://qsartoolbox.org/file/2019/02/Toolbox-4.3-Release-Notes-1.pdf).

Formatted: Font: (Asian) Korean

2. Procedure

2.1) OECD QSAR Toolbox

Note that the QSAR Toolbox operates in six consecutive flow modules starting from "Input," followed by "Profiling," "Data," "Category Definition," "Data Gap Filling," and "Report" located at the top of the program interface.

2.1.1) Explore the aforementioned six stages through six toolbar icons by left-click through your mouse pointer. Have a first look over the stages of "Input," "Data Gap Filling," and "Report" that are necessary to perform the automated workflow "Ecotoxicological endpoint" and to document

its results.

2.1.2) Take a short look over optional stages “Profiling” and “Data.” “Profiling” stage provides an initial insight into the target substance’s (eco)toxicity potential and environmental fate characteristics; optional “Data” stage enables searching for available experimental data related to the target substance.

2.2) Input

2.2.1) Upon starting the QSAR Toolbox, the user begins at the “Input” toolbox stage by default. The QSAR Toolbox creates a working file named “Document 1” automatically, which is displayed in the stage option panel on the left of the program interface. Rename the file, if desired, by right-clicking the working file.

2.2.2) Click on the “CAS#” button in the actions toolbar, enter the chemical abstract service (CAS) number of the target substance in the available text field and click “Search.” The tool then searches for the target substance by CAS number.

2.2.3) If required, choose other search options that are available in the action toolbar such as searching by substance name or simplified molecular-input line-entry system (SMILES) code. SMILES can be entered as 2D non-stereochemical or 3D stereochemical containing forms. Click buttons “Name” or “Structure,”, respectively. Use the tool “Structure” to draw the target substance.

Formatted: Highlight

2.2.4) The search tool displays the search results through database records in a pop-up window. Choose the record reporting a “high” CAS-SMILES relation (“CS Relation” field) if multiple records are retrieved for the target substance by checking the box on the left of the record. Click “OK.”

Note that to proceed from this point is possible only if the retrieved record contains a SMILES code, as the SMILES code (2D non-stereochemical containing form) is the basis for computation.

Formatted: Not Highlight

2.2.5) Batch mode: To perform the *in silico* assay for multiple target substances, write a simple substance list in a text editor in which each CAS number is listed in a single row (Supplementary Fig. S3). Save the text file with an appropriate name and extension .txt on the computer.

Formatted: Not Highlight

Formatted: Not Highlight

2.2.6) Batch mode: Click on “Data.”. Then go to “Databases” in the stage option panel on the left of the program interface. Make sure databases are checked that are listed under “Ecotoxicological Information.”

2.2.7) Batch mode: Click on “Input.”. Select “Query” from the actions toolbar. Accept the settings set in step 2.2.6 by clicking “Yes” in the dialog window.

2.2.8) Batch mode: Choose the “CAS” tab. Upload the substance list saved as text file through “Load list” from your computer.

2.2.~~8~~9) Batch mode: There are two "Add" buttons available; click the "Add" button on the bottom of the pop-up menu and then click "Execute." The QSAR Toolbox will display a message on the number of substances that have been retrieved for the search.

Note that some substances of the loaded list may not be found by the search tool or that several entries may be available for one CAS number. It is not possible to delete substances from the retrieved set of substances.

2.3) Profiling (Optional. If not required, go to 2.5)

2.3.1) Click on the toolbox stage button "Profiling." ~~-~~ Go to "Profiling methods" in the stage option panel on the left of the program interface.

2.3.2) Click "Unselect All." Check all profilers listed under "Predefined" and those related to aquatic toxicity listed under "Endpoint specific" such as "Acute aquatic toxicity classification by Verhaar (Modified)[19,20]."

Formatted: Superscript

2.3.3) Finish the selection. Then click on the "Apply" button in the actions toolbar.

Note that the QSAR Toolbox provides recommendations on a set of profilers. These are highlighted in green (suitable) and orange (plausible) when left-clicking the data matrix field next to the endpoint of interest. Available endpoints are listed in the endpoint tree next to the stage option panel.

Note that the profiler "Substance type" will indicate whether the target substance is a "discrete chemical." The information is displayed in the expanded endpoint tree "Profile," "Predefined," "Substance type." Only if the target substance is a discrete chemical can the automated workflow run successfully.

Note that "Acute aquatic toxicity classification by Verhaar (Modified)" provides a first estimate of the acute aquatic toxicity mechanism of the target substance[19,20]. The information is displayed in the expanded endpoint tree "Profile," "Endpoint Specific," "Acute aquatic toxicity classification by Verhaar (Modified)." Five classes are available: (class 1) inert chemicals (baseline toxicity); (class 2) less inert chemicals; (class 3) reactive chemicals; (class 4) specifically acting chemicals; and (class 5) for chemicals not possible to classify.

Formatted: Not Highlight

2.3.4) Click "Parameter" in the endpoint tree to run integrated 2D and 3D QSAR models available in the QSAR Toolbox, if desired. Click "Calculate/extract all parameters for all chemicals" in the pop-up menu.

2.3.5) 2D and 3D QSAR models compiled in "Parameter" provide numeric values. Use "Profiling methods" for qualitative information (see section 2.3.1).

2.4) Data (Optional. If not required, go to 2.5)

2.4.1) Click on the toolbox stage button “Data.” Then click on button “Gather” from the actions toolbar.

2.4.2) Select “All endpoints” to gather all experimental data; select “Choose” to gather endpoint specific experimental data. -As an example, if aquatic toxicity is taken as the user’s focus, the user would click on “Choose,” then on “Ecotoxicological Information,” “Aquatic toxicity,” and finally “OK.”

Note that choosing to gather experimental data for all endpoints may lead to extended processing time.

Note that the user can adapt the hierarchy of the endpoint tree to his or her own purposes. This changes the manner in which data are displayed.

2.4.3) If desired, right-click the endpoint of interest in the endpoint tree area. Choose “Set tree hierarchy” in the pop-up menu. Organize the endpoint tree in the preferred manner by using the available terms and arrows and click “OK.”

2.4.4) If desired, export the gathered data as a Microsoft Excel file. Right-click on the endpoint of interest and choose “Export Data matrix” in the pop-up menu.

2.4.5) A “Matrix export” wizard opens and enables adding other endpoints to the export list. Finish the selection, click “Export,” and save the file on the computer.

Note that exporting data from all databases is not possible. For example, data retrieved from the database “ECHA CHEM” cannot be saved.

2.5) Data gap filling

2.5.1) Click on the toolbox stage button “Data gap filling.”- Then click “Automated” in the actions toolbar.

2.5.2) Select "Ecotoxicological Endpoint,” then “Fish, LC50 (lethal concentration, 50%) at 96-h for *Pimephales promelas* (mortality). “ Click “OK.” A “Workflow Controller” appears.

2.5.3) Foresee some processing time up to several minutes, especially in batch mode.

Note that the QSAR Toolbox automatically applies a defined set of profilers when searching for suitable substances with available experimental data for the prediction. The experimental data (e.g., effect concentrations 96-h LC_{50} (*P. promelas*) or 96-h EC_{50} (*P. promelas*, mortality)) are used to generate the prediction for the target substance by either the linear approximation or nearest neighbor method. Note that the methods of linear approximation and nearest neighbor are

referred to as trend analysis (labeled as “T”) and read-across (labeled as “R”), respectively.

2.5.4) The user receives a message if the prediction is performed successfully. Click “OK” and close the workflow controller indicating “Finished workflow” by clicking “x” in the upper right corner.

2.5.5) Batch mode: Upon starting the automated workflow, the uses is asked to specify the range of substance on which to execute the workflow. Except the full range of substances selected by default in the dialog window by clicking “OK.”

2.5.6) Batch mode: The user will not receive a message indicating whether a prediction was run successfully or unsuccessfully. Close the workflow controller indicating “Finished workflow” at the end of the batch processing by clicking “x” in the upper right corner.

2.6) Report

2.6.1) If a prediction was successfully executed, click on the toolbox stage button “Report.”

Note that no reports can be generated in batch mode.

2.6.2) Scroll down and find the prediction value in matrix field located in a yellow-highlighted row next to endpoint “96-h.” The predicted value is labelled with “T” or “R.”~~.~~ Activate this specific data matrix field by left-clicking it.

2.6.3) Click the button “Prediction” in the actions toolbar. Customize the report content and appearance in the pop-up wizard. Three types of reports are available: (i) prediction, (ii) category, and (iii) data matrix.

2.6.4) The wizard allows the user to fill in the author’s name and contact details. Write a short summary, provide a detailed explanation of the mechanistic interpretation, or provide a justification for the adequacy of the prediction.

2.6.5) Include additional information related to the executed prediction, if desired. The extent of additional information depends on the user.

2.6.6) Go through the wizard by clicking “Next.” Finally, click “Create report” and save the prediction and category reports as PDF files and the data matrix as a Microsoft Excel spreadsheet on the computer.

2.6.7) Find additional details on the functionalities of the QSAR Toolbox and automated workflows in the application manual for the OECD QSAR Toolbox v. 4 (F1 help on the key board). Details on the algorithms and rationale behind the automated workflow are described by Dimitrov et al.[8] and Yordanova et al. [9]

3. Application

3.1) If the user chooses to use the predicted effect concentration (i.e., 96-h LC_{50} of *P. promelas*) in the environmental risk assessment, he or she can use the lower limit of the 95% confidence interval. Find the data on the first page of the saved prediction report (PDF) in "Prediction summary," "Predicted value: <mean> (from <lower_limit> to <upper_limit>)."

Note that the advice given here is based on results of the comparison between predicted and experimental data for a set of target substances reported in this study. Selecting the lower end of the 95% confidence range will increase the likelihood that the predicted effect concentration will not underestimate the real toxicity of the substance (please see the following section). The predicted effective concentration of the lower limit of the 95% confidence interval will therefore present a safer basis for the risk assessment.

~~**2. Procedure**~~

~~2.1) OECD QSAR Toolbox~~

~~**INFO:** The QSAR Toolbox operates in six consecutive flow modules starting from "Input," followed by "Profiling," "Data," "Category Definition," "Data Gap Filling," and "Report" located at the top of the program interface.~~

~~2.1.2) Explore the aforementioned six stages through six toolbar icons. The stages of "Input," "Data Gap Filling," and "Report" are necessary to perform automated workflows and to document results. Optional "Profiling" stage provides an initial insight into the target substance's (eco)toxicity potential and environmental fate characteristics; optional "Data" stage enables searching for available experimental data related to the target substance.~~

~~2.2) Input~~

~~2.2.1) Upon starting the QSAR Toolbox, the user begins at the "Input" toolbox stage by default. The QSAR Toolbox creates a working file named "Document 1" automatically, which is displayed in the stage option panel on the left of the program interface. Rename the file, if desired, by right-clicking the working file.~~

~~2.2.2) Click on the "CAS#" button in the actions toolbar, enter the chemical abstract service (CAS) number of the target substance in the available field and click "Search." The tool then searches for the target substance by CAS number.~~

~~**INFO:** If required, other search options are available such as searching by substance name ("Name") or simplified molecular-input line-entry system (SMILES) code ("Structure").~~

~~2.2.3) The search tool displays the search results through database records. Choose the record reporting a "high" CAS-SMILES relation ("CS Relation" field) if multiple records are retrieved for~~

~~the target substance and click "OK."~~

~~**INFO:** Note that to proceed from this point is possible only if the retrieved record contains a SMILES code, as the SMILES code is the basis for computation.~~

~~2.2.4) Batch mode: To perform the *in silico* assay for multiple target substances, write a text file in which each CAS number is listed in a single row. Save the text file.~~

~~2.2.5) Batch mode: Select "Query" from the actions toolbar, accept the default settings by clicking "Yes," chose the "CAS" tab, and upload the substance list through "Load list." There are two "Add" buttons available; click the "Add" button on the bottom of the pop-up menu and then click "Execute." The QSAR Toolbox will display a message on the number of substances that have been retrieved for the search.~~

~~**INFO:** Be aware that some substances of the loaded list may not be found by the search tool or that several entries may be available for one CAS number. It is not possible to delete substances from the retrieved set of substances. Batch processing must be applied to all retrieved substances.~~

~~2.3) Profiling (Optional. If not required, go to 2.5)~~

~~2.3.1) Click on the toolbox stage button "Profiling" and go to "Profiling methods."~~

~~2.3.2) The "Profiling methods" are listed on the left in the stage option panel. Click "Unselect All" and check all profilers listed under "Predefined" and those related to aquatic toxicity listed under "Endpoint specific" such as "Acute aquatic toxicity classification by Verhaar (Modified)." Finish the selection and click on the "Apply" button in the actions toolbar.~~

~~**TIP:** The QSAR Toolbox provides recommendations on a set of profilers. These are highlighted in green (suitable) and orange (plausible) when left-clicking the data matrix field next to the endpoint of interest. Available endpoints are listed in the endpoint tree next to the stage option panel.~~

~~**INFO:** Note that the profiler "Substance type" will indicate whether the target substance is a "discrete chemical." The information is displayed in the expanded endpoint tree "Profile," "Predefined," "Substance type." Only if the target substance is a discrete chemical can the automated workflow run successfully.~~

~~**INFO:** "Acute aquatic toxicity classification by Verhaar (Modified)" provides a first estimate of the acute aquatic toxicity of the target substance. The information is displayed in the expanded endpoint tree "Profile," "Endpoint Specific," "Acute aquatic toxicity classification by Verhaar (Modified)." Five classes are available: (class 1) inert chemicals (baseline toxicity); (class 2) less inert chemicals; (class 3) reactive chemicals; (class 4) specifically acting chemicals; and (class 5) for chemicals not possible to classify.~~

2.3.3) Right-click "Parameter" to run integrated 2D and 3D QSAR models available in the QSAR Toolbox, if desired. Click "Calculate/extract all parameters for all chemicals" in the pop-up menu.

INFO: Note that parameters provide numeric values. Use "Profiling methods" for qualitative information (see section 2.3.1).

2.4) Data (Optional. If not required, go to 2.5)

2.4.1) Click on the toolbox stage button "Data" and then button "Gather" from the actions toolbar. Select "All endpoints" to gather all experimental data; select "Choose" to gather endpoint specific experimental data. As an example, if aquatic toxicity is taken as the user's focus, the user would click on "Choose," then on "Ecotoxicological Information," "Aquatic toxicity," and finally "OK."

TIME: Be aware that choosing to gather experimental data for all endpoints may lead to extended processing time.

INFO: The user can adapt the hierarchy of the endpoint tree to his or her own purposes. This changes the manner in which data are displayed. Right-click the endpoint of interest in the endpoint tree area. Choose "Set tree hierarchy" in the pop-up menu. Organize the endpoint tree in the preferred manner and click "OK."

TIP: If desired, export the gathered data as a Microsoft Excel file. Right-click on the endpoint of interest and choose "Export Data matrix" in the pop-up menu. A "Matrix export" wizard opens and enables adding other endpoints to the export list. Finish the selection, click "Export," and save the file.

INFO: Note that exporting data from all databases is not possible. For example, data retrieved from the database "ECHA CHEM" cannot be saved.

2.5) Data gap filling

2.5.1) Click on the toolbox stage button "Data gap filling" and then on "Automated" in the actions toolbar. Select "Ecotoxicological Endpoint," then "Fish, LC50 (lethal concentration, 50%) at 96-h for *Pimephales promelas* (mortality)," and click "OK." A "Workflow Controller" appears.

TIME: This process may take up to several minutes, especially in batch mode.

INFO: The QSAR Toolbox automatically applies a defined set of profilers when searching for suitable substances with available experimental data for the prediction. The experimental data (e.g., effect concentrations 96-h LC_{50} (*P. promelas*) or 96-h EC_{50} (*P. promelas*, mortality)) are used to generate the prediction for the target substance by either the linear approximation or nearest neighbor method. Note that the methods of linear approximation and nearest neighbor are referred to as trend analysis and read-across, respectively.

2.5.2) The user receives a message if the prediction is performed successfully. Click "OK" and close the workflow controller indicating "Finished workflow."

INFO: Batch mode: The user will not receive a message indicating whether a prediction was run successfully or unsuccessfully. Close the workflow controller indicating "Finished workflow" at the end of the batch processing.

2.6) Report

2.6.1) If a prediction was successfully executed, click on the toolbox stage button "Report."

INFO: Note that no reports can be generated in batch mode.

2.6.2) Scroll down and find the relevant data matrix field located in a yellow-highlighted row next to endpoint "96-h." Activate this specific data matrix field by left-clicking it.

2.6.3) Click on the button "Prediction" in the actions toolbar and, in the pop-up wizard, customize the report content and appearance. Three types of reports are available: (i) prediction, (ii) category, and (iii) data matrix.

INFO: The wizard allows the user to fill in the author's name and contact details. Write a short summary, provide a detailed explanation of the mechanistic interpretation used, or provide a justification for the adequacy of the prediction. Include additional information related to the executed prediction, if desired. The extent of additional information depends on the user.

2.6.4) In the wizard, click "Next." Finally, click "Create report" and save the prediction and category reports as PDF files and the data matrix as a Microsoft Excel spreadsheet.

TIP: Additional details on the functionalities of the QSAR Toolbox and automated workflows can be found in the application manual for the OECD QSAR Toolbox v. 4 (F1 help on the key board). Details on the algorithms and rationale behind the automated workflow are described by Dimitrov et al.[8] and Yordanova et al.[9]

3. Application

3.1) If the user chooses to use the predicted effect concentration (i.e., 96-h LC_{50} of *P. promelas*) in the environmental risk assessment, he or she can use the lower limit of the 95% confidence interval. Find the data on the first page of the saved prediction report (PDF) in "Prediction summary," "Predicted value: <mean> (from <lower_limit> to <upper_limit>)."

INFO: The advice given here is based on results of the comparison between predicted and experimental data for a set of target substances reported in this study. Selecting the lower end of the 95% confidence range will increase the likelihood that the predicted effect concentration will not underestimate the real toxicity of the substance (please see the following section). The

~~predicted effective concentration of the lower limit of the 95% confidence interval will therefore present a safer basis for the risk assessment.~~

REPRESENTATIVE RESULTS:

The example protocol described in this study was implemented for quantitative analysis and prediction of acute toxicities of selected EDs in fish. ~~Based on assessments of the entire species available from the database, values for the predicted and experimental 96-h $\log_{10}LC_{50}$ exhibited linearity with the $\log_{10}K_{OW}$ values in the domain between -1 and 7, indicating a hyperbolic correlation between LC_{50} and K_{OW}. An overall trend existed whereby the LC_{50} decreased for higher K_{OW} values of EDs for the data obtained from both computational predictions and experiments, suggesting increasing acute toxicity in fish species for EDs with higher hydrophobicity. Fig. 3 and the following Eqs. 1 and 2 show the statistical relationship between the descriptor (K_{OW}) and endpoint (LC_{50}) of the EDs listed in Table 1. In the log-log plot, the linear regression for the predicted and experimental values showed a comparable slope ($\log_{10}LC_{50}$ / $\log_{10}K_{OW}$ = -0.508 and -0.411, respectively) with a relatively high coefficient of determination for the prediction (adjusted r^2 = 0.807 and 0.446 for prediction and observation, respectively).~~

~~*Computational prediction:* $\log_{10}LC_{50} = -0.508 \cdot \log_{10}K_{ow} + 2.43$ (Eq. 1)~~

~~*Experimental observation:* $\log_{10}LC_{50} = -0.411 \cdot \log_{10}K_{ow} + 2.11$ (Eq. 2)~~

When the predicted data points were plotted versus experimental data points as a log-log scale, a positive correlation between both was found for all fish and for a representative species, namely, *Pimephales promelas* (fathead minnow) (Fig. 34). In both cases, the slope of the linear regression appeared to be comparable (Predicted LC_{50} / Experimental LC_{50} = 0.61125 and 0.60213 for all fish and *P. promelas*, respectively). Because of the limited amount of experimental data, the number of available values from experimental observation was usually smaller than that from computational prediction. Application of the tolerance factor as 5-folds for the computational capability[2119] resulted in 945% (346/368) and 96% (26/27) of the protective prediction for all fish and *P. promelas*, respectively. Based on this prediction, 3',5,7-trihydroxy-4',6-dimethoxyisoflavone and 1,4-benzenediol appeared to exhibit calculated LC_{50} values greater than the tolerance limit.

To enable safety assessment at the highest reliability, further computational analysis was performed by plotting the predicted lower limit of the 95% confidence interval of LC_{50} (instead of the mean values used in Fig. 32) versus the experimentally derived values (Fig. 45). In this evaluation with an elevated safety threshold, 92% (335/368) of the total tested endocrine disrupting compounds were shown to fall into the protective range when compared to the experimentally derived values except for 3',5,7-trihydroxy-4',6-dimethoxyisoflavone, 1,4-benzenediol, and 4-hexylphenol.

Based on assessments of the entire species available from the database, values for the predicted and experimental 96-h $\log_{10}LC_{50}$ exhibited linearity with the $\log_{10}K_{OW}$ values in the domain between -1 and 7, indicating a hyperbolic correlation between LC_{50} and K_{OW}. An overall trend existed whereby the LC_{50} decreased for higher K_{OW} values of EDs for the data obtained from both

Formatted: Left

computational predictions and experiments, suggesting increasing acute toxicity in fish species for EDs with higher hydrophobicity (Supplementary Fig. S1).

By the rule-based ER profiler embedded in the OECD QSAR Toolbox, the ER binding affinities of the EDs were categorized as non-, weak, moderate, strong, and very strong binders in order of increasing binding affinity[18]. Accordingly, the statistical distribution of $\log_{10}K_{ow}$ could be displayed as a qualitative classification of ER binding affinity (Supplementary Fig. ~~6~~S2). Overall, the changes in K_{ow} distribution ranges and their mean levels appeared not to have a defined tendency. Similarly, the distributions of predicted and experimental LC_{50} were shown as the extent of the ER binding affinity (Fig. 5~~7~~). In this case, the mean levels of predicted LC_{50} for ER binders were higher than those of non-binders. By contrast, for the experimental LC_{50}, the mean levels of non- and weak binders were higher than those of stronger ER binders.

FIGURE AND TABLE LEGENDS:

Table 1: List of evaluated endocrine disrupting chemicals. Average mean (AVE) and lower 95% confidence interval (CI) effective concentrations (95-h LC_{50}, *Pimephales promelas*) as well as Estrogen Receptor Binding were predicted with the QSAR Toolbox version 4.3 Automated Workflow. $Log_{10}K_{ow}$ was retrieved via QSAR Toolbox version 4.3 from KOWWIN v1.68, 2000, U.S. Environmental Protection Agency. Experimental $log_{10}K_{ow}$ values were preferred over predicted values. The target substance list was compiled from previously reported lists of EDs[2~~20~~-2~~12~~].

Figure 1: Basic scheme of the general workflow of the OECD QSAR Toolbox.

Figure 2: Flow conceptualizing the modules and sequences applied to predict the acute toxicities of endocrine disruptors (EDs) in fish using the OECD QSAR Toolbox.

~~**Figure 3:** Relationships between 96-h LC_{50} and $log_{10}K_{OW}$ values of EDs listed in Table 1. For all fish species, the values for predicted (black empty squares, n = 77) and experimental (red empty circles, n = 38) are plotted together in a log scale as a function of $log_{10}K_{OW}$. For the predicted LC_{50}, the average values are displayed. The gray (Eq. 1) and light-red (Eq. 2) lines represent the linear regression for each case (predicted and experimental, respectively).~~

Figure 3~~4~~: Predicted versus experimental 96-h LC_{50} of EDs in Table 1 for all fish (blue diamonds, n = 3~~6~~8) and for a selected species, *P. promelas* (cyan diamonds, n = 27). For the predicted LC_{50}, the average ("AVE") values are displayed. The dashed lines represent the linear regressions for the two groups: for all fish (light blue), $Predicted\ LC_{50}^{AVE} = 0.611~~25~~ \cdot (Experimental\ LC_{50}) + 0.277~~89~~$ (adjusted $r^2 = 0.40~~1~~8$), and for *P. promelas* (light cyan), $Predicted\ LC_{50}^{AVE} = 0.602~~13~~ \cdot (Experimental\ LC_{50}) + 0.385~~98~~$ (adjusted $r^2 = 0.4152$). The solid diagonal line shows unity in which the predicted and experimental values are equal[2~~19~~]. The dotted gray line shows the 5-fold tolerance limit of the computational capability[19]. Outliers: 3',5,7-trihydroxy-4',6-dimethoxyisoflavone (*) and 1,4-benzenediol (**).

Figure 4~~5~~: Predicted (lower limit of the 95% confidence interval; "low-95%") versus

experimental 96-h LC_{50} of EDs in Table 1 for all fish (n = 368). The dashed line represents the linear regression: $Predicted\ LC_{50}^{AVE} = 0.47093 \cdot (Experimental\ LC_{50}) - 0.312293$, where adjusted r^2 = 0.193207. The solid diagonal line indicates unity where the predicted and experimental values are equal to each other[19]. Outliers: 3',5,7-trihydroxy-4',6-dimethoxyisoflavone (*), 1,4-benzenediol (**), and 4-hexylphenol (***).

~~**Figure 6:** Distributions of the $log_{10}K_{OW}$ depending on the estrogen receptor (ER) binding affinity of EDs in Table 1 for all fish (n = 8-22 for each category). A box plot represents: (a) mean (small square with a horizontal bar), (b) 1st and 3rd quartiles (lower and upper ends of the box, respectively), (c) median (horizontal segment inside the box), (d) 5th and 95th percentile (lower and upper error bars, respectively), (e) 1th and 99th percentile (lower and upper ×, respectively), and (f) minimum and maximum (lower and upper –, respectively).~~

Figure 57: Distributions of the predicted (solid boxes; n = 8-202 for each category) and experimental (dashed boxes; n = 3-16 for each category) 96-h LC_{50} depending on the ER binding affinity of EDs in Table 1 for all fish. A box plot represents: (a) mean (small square with a horizontal bar), (b) 1st and 3rd quartiles (lower and upper ends of the box, respectively), (c) median (horizontal segment inside the box), (d) 5th and 95th percentile (lower and upper error bars, respectively), (e) 1th and 99th percentile (lower and upper ×, respectively), and (f) minimum and maximum (lower and upper -, respectively).

DISCUSSION:

The versatility of the OECD QSAR Toolbox as analytic software for ecotoxicology was shown here with specific interest in the adverse effects of endocrine disrupting chemicals on aquatic vertebrates. In addition, a simple and standard protocol was demonstrated for predicting acute toxicity (96-h LC_{50}) of 747 representative EDs (Table 1) for fish species. This was achieved by applying category building, data gap filling, and ER profiling modules embedded in the QSAR Toolbox (Figs. 1 and 2).

The linear correlation between $log_{10}LC_{50}$ and $log_{10}K_{OW}$ with a negative slope (as shown in Supplementaray Fig. S13) has long been known as a standard quantitative relationship in QSAR analyses[253], where higher toxicity is shown the more hydrophobic a given chemical is. As can be seen from a simple calculation, the general mathematical relation that includes Eqs. 1 and 2,

$$log(LC_{50}) = a' - b \cdot log(K_{ow}) \qquad \text{(Eq. 3)}$$

is a converted expression from the following power function[264]:

$$LC_{50} = a \cdot {K_{ow}}^{-b}, \text{ where } a' = log(a) \qquad \text{(Eq. 4)}$$

From the plot of (4), characterizing an intermediate range of K_{OW}[264] may be possible by adjusting the parameters a and b, where a certain variation in hydrophobicity (or hydrophilicity) does not significantly change the endpoint of acute toxicity.

Comparative analyses between the computational predictions and experimental observations on the LC_{50}, as shown in Figs. 34 and 45, have been typically reported in studies of QSAR for various aquatic toxicants, including technical nonionic surfactants[275], triazole fungicides[286], and pesticide metabolites[2119]. This type of retrospective validation provides information on how far

a given QSAR tool can reach in terms of comparative performance to experimental results. In our study of acute toxicity in fish, the QSAR Toolbox is proven to provide protective predictions for over 90% of the tested EDs on all fish and on a single species, *Pimephales promelas*.

Further identifying the three outlier chemicals in Figs. 34 and 45, which showed higher predicted LC_{50} on average and at a minimum, respectively, is required. First, the 3',5,7-trihydroxy-4',6-dimethoxyisoflavone is a type of flavonoid (more specifically, an isoflavone), which is considered to be generally safe and is used in herbal pharmaceuticals but still has estrogen-related concerns[29~~7~~]~~.~~ and may cause acute toxicity probably through oxidative phosphorylation uncoupling[30]. Next, the 1,4-benzenediol, called hydroquinone, is a phenolic compound that can trigger a non-specific and cytotoxic immune response in fish[31~~28~~]. Finally, the 4-hexylphenol has been known to exhibit sufficient positive estrogenic activity to be classified as an ED[32~~29~~]. It is well studied~~known~~ that the main reason of the acute toxicity of hydroquinone is the reduction-oxidation (redox) cycling: The hydroquinone is oxidized to benzoquinone and reduced back to semi-quinone or hydroquinone repeatedly, with depleting cofactors and generating reactive oxygen species[33]. ~~All of these three~~ The other two chemicals may require deeper investigations to reveal their mechanisms of action in acute ecotoxicity using molecular docking approaches such as that of Panche et al.[34~~0~~], which cannot be covered by the QSAR Toolbox. ~~At this point, applying 3D-QSAR modeling is a necessity[3].~~

EDs interfere with the endocrine system mainly through physicochemical interactions with steroid receptors such as estrogen and androgen receptors, which are of considerable interest in QSAR modeling studies[35~~1~~]. Considering this, the QSAR Toolbox is robust in terms of facile and rapid classification of ER binding affinities for a set of chemicals based only on the 2D descriptors of molecular structures~~, such as SMILES codes[18]~~. When this ER profiler system was applied to our list of EDs, we first found no clear correlation between ER binding affinity and hydrophobicity (Supplementary Fig. S2~~6~~). This result might be explained by the fact that the formation of a steroid-receptor complex is not a direct consequence of a hydrophobic bonding contribution but should be accompanied by a conformational change in the active-site receptor structure[36~~2~~]. The receptor binding can be also due to hydrogen-bonding and π-stacking. Additionally, the position of each chemical group on the molecule may affect the receptor binding, even if the hydrophobicity and number of hydrogen-bond acceptors-donors remain the same. Second, the ER profiler produced contrary trends between predicted and experimental LC_{50} mean levels with increasing ER binding affinity (Fig. 5~~7~~). This may be because the lethality of parents in an acute toxicity test would not be due to ER binding but rather to narcosis in most cases, or to redox cycling in the case of hydroquinone, for example. ~~This suggests a fundamental predictive limitation of the conventional QSAR tools, where there is a lack of multi-step pathway linkages between the molecular initiating event (ER binding of EDs) and the adverse outcome (endpoint, including LC_{50})[33]. However,~~ more extensive analysis, including the chronic toxicity, is required on a larger set of EDs to define the predictive limitations of the current version of the QSAR Toolbox.

This preliminary research might also have public health implications because the steroids (androgens, estrogens, progestines, and corticoids) and their receptors exhibit similar or even

Formatted: Superscript

Formatted: Superscript

Formatted: Font: (Asian) Korean

identical macromolecular structures across vertebrates[5]. These types of analogous endocrine signaling systems may operate using a common mechanism in key events of EDs[5] ~~(Hayes, 2005)~~. Nevertheless, additional and complementary methodologies are required to illuminate this vast and complex aspect, for example, by performing computational modeling of absorption, distribution, metabolism, and excretion (ADME), and/or adverse outcome pathway (AOP)[3,4]. Furthermore, because most of the scientific and public concerns raised about the adverse effects of EDs are related to their chronic toxicities, improving the databases and algorithms in the QSAR Toolbox and producing reliable long-term ecotoxicology predictions for EDs are both necessary.

In this paper, we have demonstrated the application of QSAR Toolbox to compare ecotoxicological LC_{50} values for fish with $\log_{10}K_{ow}$ values of EDs. Throughout the suggested protocol, it results in little relationship between the two parameters, as have been revealed by previous studies (Kim et al.,[39] for example) that $\log_{10}K_{ow}$ is not a good direct predictor of aquatic LC_{50}. ~~However, most of the current QSARs use log10Kow as explanatory variable (Melnikov et al. 2016), as it is difficult to find other more suitable but simple input parameter with predictive power (#reference).~~ In spite of this limitation, our protocol provides a general review or "vignette" to describe how to use the dashboard for a given purpose, since it is a valid application to use the QSAR Toolbox for investigating correlation between LC_{50} (or ER binding affinity) and $\log_{10}K_{ow}$. Nevertheless, it should be noted that (1) illuminating the link between estrogen receptor binding and chronic toxicity, rather than acute toxicity (lethality), would be more relevant so that clearer correlation(s) might be found, and (2) the androgen receptor, together with that of estrogen, also plays a critical role in reproductive toxicity. Therefore, it is required for the future version of the QSAR Toolbox to improve the prediction functions in light of those two points.

ACKNOWLEDGMENTS:

This research was supported by the National Research Council of Science & Technology (NST) grant by the South Korean government (MSIP) (No. CAP-17-01-KIST Europe) and Project 11911.

DISCLOSURES:

The authors have nothing to disclose.

REFERENCES:

1 Najarian, K., Najarian, S., Gharibzadeh, S., Eichelberger, C. N. *Systems Biology and Bioinformatics: A Computational Approach.* CRC Press. Boca Raton, FL, USA (2009).

2 Fujita, T., Iwasa, J., Hansch, C. A new substituent constant, π, derived from partition coefficient. *J. Am. Chem. Soc.* **86**, 5175-5180 (1964).

3 Roy, K., Kar, S., Das, R. N. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment.* Academic Press. Cambridge, MA, USA (2015).

4 Raies, A. B., Bajic, V. B. In silico toxicology: computational methods for the prediction of chemical toxicity. *WIREs Comput. Mol. Sci.* **6**, 147-172 (2016).

5 Hayes, T. B. Welcome to the revolution: integrative biology and assessing the impact of

Formatted: Superscript
Formatted: Highlight
Formatted: Subscript
Formatted: Subscript
Formatted: Subscript
Formatted: Superscript
Formatted: Subscript
Formatted: Subscript
Formatted: Subscript
Formatted: Subscript
Formatted: Subscript
Formatted: Subscript
Formatted: Font: (Asian) Korean

endocrine disruptors on environmental and public health. *Integr. Comp. Biol.* **45**, 321-329 (2005).

6 Schug, T. T., Johnson, A. F., Birnbaum, L. S., Colborn, T., Guillette, L. J. Jr., Crews, D. P., Collins, T., Soto, A. M., Vom Saal, F. S., McLachlan, J. A., Sonnenschein, C., Heindel, J. J. Minireview: endocrine disruptors: past lessons and future directions. *Mol. Endocrinol.* **30**, 833-847 (2016).

7 Devillers, J., Marchand-Geneste, N., Carpy, A., Porcher, J. M. SAR and QSAR modeling of endocrine disruptors. *SAR QSAR Environ. Res.* **17**, 393-412 (2006).

8 Dimitrov, S. D., Diderich, R., Sobanski, T., Pavlov, T. S., Chankov, G. V., Chapkanov, A. S., Karakolev, Y. H., Temelkov, S. G., Vasilev, R. A., Gerova, K. D., Kuseva, C. D., Todorova, N. D., Mehmed, A. M., Rasenberg, M., Mekenyan, O. G. QSAR Toolbox - workflow and major functionalities. *SAR QSAR Environ. Res.* **27**, 203-219 (2016).

9 Yordanova, D., Schultz, T. W., Kuseva, C, Tankova, K., Ivanova, H., Dermen, I., Pavlov, T., Temelkov, S., Chapkanov, A., Georgiev, M., Gissi, A., Sobanski, T., Mekenyan, O. G. Automated and standardized workflows in the OECD QSAR Toolbox. *Comput. Toxicol.* **10**, 89-104 (2019).

10 Mombelli, E., Devillers, J. Evaluation of the OECD (Q)SAR Application Toolbox and Toxtree for predicting and profiling the carcinogenic potential of chemicals. *SAR QSAR Environ. Res.* **21**, 731-752 (2010).

11 Devillers, J., Mombelli, E., Samsera, R. Structural alerts for estimating the carcinogenicity of pesticides and biocides. *SAR QSAR Environ. Res.* **22**, 89-106 (2011).

12 Li, C., Wang, D., Li, N., Luo, Q., Xu, X., Wang, Z. Identifying unknown by-products in drinking water using comprehensive two-dimensional gas chromatography-quadrupole mass spectrometry and in silico toxicity assessment. *Chemosphere* **163**, 535-543 (2016).

13 Devillers, J., Mombelli, E. Evaluation of the OECD QSAR Application Toolbox and Toxtree for estimating the mutagenicity of chemicals. Part 1. Aromatic amines. *SAR QSAR Environ. Res.* **21**, 753-769 (2010a).

14 Devillers, J., Mombelli, E. Evaluation of the OECD QSAR Application Toolbox and Toxtree for estimating the mutagenicity of chemicals. Part 2. α-β unsaturated aliphatic aldehydes. *SAR QSAR Environ. Res.* **21**, 771-783 (2010b).

15 Kulkarni, S. A., Barton-Maclaren, T. S. Performance of (Q)SAR models for predicting Ames mutagenicity of aryl azo and benzidine based compounds. *J. Environ. Sci. Health C Environ. Carcinog. Ecotoxicol. Rev.* **32**, 46-82 (2014).

16 Craig, E. A., Wang, N. C., Zhao, Q. J. Using quantitative structure-activity relationship modeling to quantitatively predict the developmental toxicity of halogenated azole compounds. *J. Appl. Toxicol.* **34**, 787-794 (2014).

17 Tebby, C., Mombelli, E., Pandard, P., Péry, A. R. Exploring an ecotoxicity database with the OECD (Q)SAR Toolbox and DRAGON descriptors in order to prioritise testing on algae, daphnids, and fish. *Sci. Total Environ.* **409**, 3334-3343 (2011).

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

19 Verhaar, H. J. M., van Leeuwen, C. J., Hermens, J. L. M. Classifying environmental pollutants. 1: structure-activity relationships for prediction of aquatic toxicology. *Chemosphere* **25**, 471-491 (1992).

Formatted: French (France)

Formatted: French (France)

20 Enoch, S. J., Hewitt, M., Cronin, M. T. D., Azam, S., Madden, J. C. Classification of chemicals according to mechanism of aquatic toxicity: an evaluation of the implementation of the Verhaar scheme in Toxtree. *Chemosphere* **73**, 243-248 (2008).

21~~19~~ Burden, N., Maynard, S. K., Weltje, L., Wheeler, J. R. The utility of QSARs in predicting acute fish toxicity of pesticide metabolites: a retrospective validation approach. *Regul. Toxicol. Pharmacol.* **80**, 241-246 (2016).

22~~0~~ Nendza, M., Wenzel, A., Müller, M., Lewin, G., Simetska, N., Stock, F., Arning, J. Screening for potential endocrine disruptors in fish: evidence from structural alerts and in vitro and in vivo toxicological assays. *Environ. Sci. Eur.* **28**, 26 (2016).

23~~1~~ Roncaglioni, A., Piclin, N., Pintore, M., Benfenati, E. Binary classification models for endocrine disrupter effects mediated through the estrogen receptor. *SAR QSAR Environ. Res.* **19**, 697-733 (2008).

24~~2~~ Sosnovcová, J., Rucki, M., Bendová, H. Estrogen receptor binding affinity of food contact material components estimated by QSAR. *Cent. Eur. J. Public Health* **24**, 241-244 (2016).

25~~3~~ Walker, J. D., Dearden, J. C., Schultz, T. W., Jaworska, J., Comber, M. H. I. QSARs for New Practitioners. In: Walker, J. D. (ed.) *QSARs for Pollution Prevention, Toxicity Screening, Risk Assessment, and Web Applications*. SETAC Press. Pensacola, FL, USA (2003).

26~~4~~ Sánchez-Bayo, F. From simple toxicological models to prediction of toxic effects in time. *Ecotoxicology* **18**, 343-354 (2009).

27~~5~~ Sjöström, M., Lindgren, Å., Uppgård, L-L. Joint Multivariate Quantitative Structure-Property and Structure-Activity Relationships for a Series of Technical Nonionic Surfactants. In: Chen, F., Schüürmann, G. (eds.) *Quantitative Structure-Activity Relationships in Environmental Sciences-VII.* SETAC Press. Pensacola, FL, USA (1997).

28~~6~~ Ding, F., Guo, J., Song, W., Hu, W., Li, Z. Comparative quantitative structure-activity relationship (QSAR) study on acute toxicity of triazole fungicides to zebrafish. *Chemistry Ecology* **27**, 359-368 (2011).

29~~7~~ Galati, G., O'Brien, P. J. Potential toxicity of flavonoids and other dietary phenolics: significance for their chemopreventive and anticancer properties. *Free Radic. Biol. Med.* **37**, 287-303 (2004).

30 Russom, C. L., Bradbury, S. P., Broderius, S. J. Predicting modes of action from chemical structure: acute toxicity in the fathead minnow (Pimephales promelas). *Environ. Toxicol. Chem.* **16**, 948-967 (1997).

31~~28~~ Taysse, L., Troutaud, D., Khan, N. A., Deschaux, P. Structure-activity relationship of phenolic compounds (phenol, pyrocatechol and hydroquinone) on natural lymphocytotoxicity of carp (*Cyprinus carpio*). *Toxicology* **98**, 207-214 (1995).

32~~29~~ Nishihara, T., Nishikawa, J., Kanayama, T., Dakeyama, F., Saito, K., Imagawa, M., Takatori, S., Kitagawa, Y., Hori, S., Utsumi, H. Estrogenic activities of 517 chemicals by yeast two-hybrid assay. *J. Health Sci.* **46**, 282-298 (2000).

33 Bolton, J. L., Trush, M. A., Penning, T. M., Dryhurst, G., Monks, T. J. Role of quinones in toxicology. *Chem. Res. Toxicol.* **13**, 135-160 (2000).

34~~0~~ Panche, A. N., Diwan, A. D., Chandra, S. R. Flavonoids: an overview. *J. Nutrit. Sci.* **5**, e47

Formatted: German (Germany)

(2016).
351 Li, J., Gramatica, P. QSAR classification of estrogen receptor binders and pre-screening of potential pleiotropic EDCs. *SAR QSAR Environ. Res.* **21**, 657-669 (2010).
362 Bohl, M. *Molecular Structure and Biological Activity of Steroids*. CRC Press. Boca Raton, FL, USA (2017).
373 Kaminuma, T., Takai-Igarashi, T., Nakano, T., Nakata, K. Modeling of signaling pathways for endocrine disruptors. *BioSystems* **55**, 23-31 (2000).
384 Lillicrap, A., Belanger, S., Burden, N., Pasquier, D. D., Embry, M. R., Halder, M., Lampi, M. A., Lee, L., Norberg-King, T., Rattner, B. A., Schirmer, K., Thomas, P. Alternative approaches to vertebrate ecotoxicity tests in the 21st century: a review of developments over the last 2 decades and current status. *Environ. Toxicol. Chem.* **35**, 2637-2646 (2016).
~~Russom, C. L., Bradbury, S. P., Broderius, S. J. Predicting modes of action from chemical structure: acute toxicity in the fathead minnow (Pimephales promelas). *Environ. Toxicol. Chem.* **16**, 948-967 (1997).~~
~~Bolton, J. L., Trush, M. A., Penning, T. M., Dryhurst, G., Monks, T. J. Role of quinones in toxicology. *Chem. Res. Toxicol.* **13**, 135-160 (2000).~~
39 Kim, J.-W., Ishibashi, H., Yamauchi, R., Ichikawa, N., Takao, Y., Hirano, M., Koga, M., Arizono, K. Acute toxicity of pharmaceutical and personal care products on freshwater crustacean (Thamnocephalus platyurus) and fish (Oryzias latipes). *J. Toxicol. Sci.* **34**, 227-232 (2009).
~~Verhaar, H. J. M., van Leeuwen, C. J., Hermens, J. L. M. Classifying environmental pollutants. 1: structure-activity relationships for prediction of aquatic toxicology. *Chemosphere* **25**, 471-491 (1992).~~
~~Enoch, S. J., Hewitt, M., Cronin, M. T. D., Azam, S., Madden, J. C. Classification of chemicals according to mechanism of aquatic toxicity: an evaluation of the implementation of the Verhaar scheme in Toxtree. *Chemosphere* **73**, 243-248 (2008).~~

Formatted: Font: Italic
Formatted: Font: Bold
Formatted: Font: Italic
Formatted: Font: Bold
Formatted: Font: Italic
Formatted: Font: Bold
Formatted: English (United States)
Formatted: Font: Italic
Formatted: Font: Bold
Formatted: Font: Italic
Formatted: Font: Bold