Received: 29 January 2021 | Revised: 13 April 2021

DOI: 10.1002/jcc.26551

SOFTWARE NOTE

Journal of COMPUTATIONAL CHEMISTRY WILEY

QSARINS-Chem standalone version: A new platform-independent software to profile chemicals for physico-chemical properties, fate, and toxicity

Nicola Chirico[1] | Alessandro Sangion[1,2] | Paola Gramatica[1] | Linda Bertato[1] | Ilaria Casartelli[1] | Ester Papa[1]

[1]Department of Theoretical and Applied Sciences, University of Insubria, Varese, Italy

[2]Department of Physical and Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario, Canada

Correspondence
Ester Papa and Paola Gramatica, Department of Theoretical and Applied Sciences, University of Insubria, via J.H. Dunant 3, 21100 Varese, Italy.
Email: ester.papa@uninsubria.it (E. P.) and paola.gramatica@uninsubria.it (P. G.)

Funding information
European Chemical Industry Council, Grant/Award Number: CEFIC-LRI ECO44; PhD Course in Chemical and Environmental Sciences (DISCA - University of Insubria), Grant/Award Number: PhD fellowship

Abstract

The new software QSARINS-Chem standalone version is a multiplatform tool, freely downloadable, for the *in silico* profiling of multiple properties and activities of organic chemicals. This software, which is based on the concept of the QSARINS-chem module embedded in the QSARINS software, has been fully redesigned and redeveloped in the Java™ language. In addition to a selection of models included in the old module, the new software predicts biotransformation rates and aquatic toxicities of pharmaceuticals and personal care products in multiple organisms, and offers a suite of tools for the analysis of predictions. Furthermore, a comprehensive and transparent database of molecular structures is provided. The new QSARINS-Chem standalone version is an informative and solid tool, which is useful to support the assessment of the potential hazard and risks related to organic chemicals and is dedicated to users which are interested in the application of QSARs to generate reliable predictions.

KEYWORDS

alternatives to animal testing, *in silico* predictions, QSAR, QSARINS, virtual screening

1 | INTRODUCTION

Chemical pollution has great impact on human and environmental health and the development of strategies to guarantee a more sustainable use of chemicals is a main challenge for chemical regulations worldwide.[1–3] The need to properly address and manage chemical risks as well as to track and substitute potentially hazardous chemicals with less dangerous ones, has in the last decade pushed toward a faster development and integration of *in vitro* and *in silico* strategies within regulations. The effort spent in traditional and regulatory science to facilitate the application of *in silico* tools, making them more transparent, easy to apply, and efficient, is major.[3] *In silico* approaches, such as models based on Quantitative Structure Activity Relationships (QSAR), are used to predict many different properties and activities of regulatory interest and for different chemical categories.[4–17] These models are useful to fill data gaps, for virtual screenings, and/or for the identification of safer alternatives to unsafe pollutants. Furthermore, the availability of multiple models, which can be combined to generate consensus predictions, helps to reduce the uncertainty associated with the prediction of a single property/activity, they can cross-validate *in silico* predictions and experiments, and support decision making processes.[4–13]

QSARINS-Chem[17] was proposed in 2014 as an additional module embedded in the software QSARINS[18] to provide a database to store models and a tool to facilitate their application. The QSARINS-Chem module included a database of chemical structures (with 3D

TABLE 1 List of the datasets[7–9,22–48] included in the structural database of QSARINS-Chem standalone version

Class[a]	Endpoint-type	Dataset name
1. General	1. Physico-Chemical properties	1. Soil organic carbon-water partition coefficient (K_{OC})[22]
	2. Environmental Persistence	1. Sediment half-lives[23]
		2. Soil half-lives[23]
		3. Water half-lives[23]
		4. Air half-lives[23]
		5. NO_3 reactivity[24]
		6. O_3 reactivity[25]
		7. OH reactivity[26]
		8. Global half-life index (GHLI)[23]
	3. Bioconcentration Factor (BCF)	1. BCF-Fernandez[9]
		2. BCF-Lu[27]
	4. Metabolic Transformation	1. Fish biotransformation[28–31]
		2. Human biotransformation Model 1[32]
		3. Human biotransformation Model 2[32]
		4. Human biotransformation Model 3[32]
		5. Human biotransformation Model 4[32]
		6. Human total elimination[32]
	5. Aquatic Toxicity	1. Fish acute toxicity (*P. promelas*)[33]
	6. Endocrine Disruption	1. Estrogen receptor binding[34]
2. Aromatic Amines	1. Mutagenicity	1. Aromatic Amines mutagenicity TA98[35]
		2. Aromatic amines mutagenicity TA100[35]
3. (Benzo)Triazoles	1. Physico-Chemical properties	1. (B)TAZ Kow[36]
		2. (B)TAZ solubility in water[36]
		3. (B)TAZ vapor pressure[36]
		4. (B)TAZ melting point[36]
	2. Aquatic Toxicity	1. (B)TAZ Algae acute toxicity (*P. subcapitata*)[37]
		2. (B)TAZ *Daphnia sp* acute toxicity[38]
		3. (B)TAZ fish acute toxicity (*O. mykiss*)[38]
4. Brominated flame retardants (BFR)	1. Physico-Chemical properties	1. BFR Kow[39]
		2. BFR Koa[39]
		3. BFR vapor pressure[39]
		4. BFR solubility in water[39]
		5. BFR Henry law constant[39]
		6. BFR melting point[39]
	2. Endocrine Disruption	1. BFR-DR-Ag[40]
		2. BFR-ER-Ag[40]
		3. BFR-ERODind[40]
		4. BFR-PR-ant[40]
		5. BFR-SULT-REP[40]
		6. BFR-T4-REP[40]
		7. BFR receptor binding affiniy[40]
5. Dioxin analogues	1. Biochemical Activity	1. Dioxin analogues pAH[41]
		2. Dioxin analogues pRB[41]
6. Esters	1. Physico-Chemical properties	1. Esters flash point[42]
	2. Aquatic Toxicity	1. Esters Algae acute toxicity[43]

(Continues)

TABLE 1 (Continued)

Class[a]	Endpoint-type	Dataset name
		2. Esters *Daphnia sp* acute toxicity[43] 3. Esters fish acute toxicity (*P. promelas*)[43] 4. Esters aquatic toxicity index (EATIN)[43]
7. Fragrances	1. Terrestrial Toxicity	1. Fragrances oral toxicity (Rat)[44]
	2. Biochemical Activity	1. Fragrances inhibition NADHox[44] 2. Fragrances Mitochondrial memb pot[44]
8. Nitrated polycyclic aromatic hydrocarbons (PAH)	1. Mutagenicity	1. NitroPAH mutagenicity TA100[45]
9. Perfluorinated compounds	1. Physico-Chemical properties	1. PFC critical Micelle concentration[46] 2. PFC solubility in water[46] 3. PFC vapor pressure[46]
	2. Terrestrial Toxicity	1. PFC oral toxicity (Rat)[47] 2. PFC oral toxicity (Mouse)[47] 3. PFC inhalation toxicity (Rat)[48] 4. PFC inhalation toxicity (Mouse)[48]
10. Personal care products	1. Aquatic Toxicity	1. PCP Algae acute toxicity (*P. subcapitata*)[7] 2. PCP *Daphnia sp* acute toxicity[7] 3. PCP fish acute toxicity (*P. promelas*)[7]
11. Pharmaceuticals	1. Aquatic Toxicity	1. Pharm. Algae acute toxicity (*P. subcapitata*)[8] 2. Pharm. *Daphnia sp* acute toxicity[8] 3. Pharm. fish acute toxicity (*O.mykiss*)[8] 4. Pharm. fish acute toxicity (*P. promelas*)[8]

[a]Class 1 includes heterogeneous structures; Class 2–11 refer to specific chemical classes or classes of use.

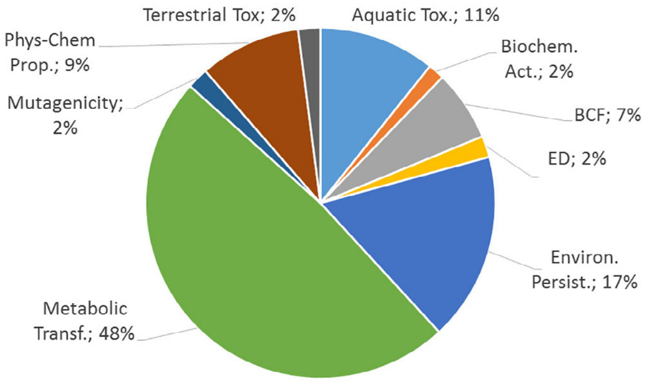

FIGURE 1 Relative abundance of records for each endpoint type: Aquatic Toxicity (Aquatic Tox.); biochemical activity (Biochem. Act.); Bioconcentration Factor (BCF); Endocrine Disruption (ED); Environmental Persistence (Environ. Persist.); Metabolic Transformation (Metabolic Transf.); Mutagenicity; Physico-Chemical properties (Phys-Chem prop.); Terrestrial Toxicity (Terrestrial Tox)

representation) and experimental data for different end-points (physico-chemical properties and biological activities) in addition to multiple linear regression (MLR) models based on descriptors calculated by the free software PaDEL-Descriptor.[19] The QSAR model reporting format (QMRF) documents, which show the compliance of the models with the "OECD principles for the validation, for regulatory purposes, of (Q)SARs"[20], were also included in the QSARINS-Chem module.[17] Being part of the QSARINS software, this module was developed in C ++ language and was available, under license agreement, only for Windows™ operative system.

In this paper we introduce the QSARINS-Chem standalone version, a new software developed in Java™ language to run on multiple platforms. This software, based on the concept of the QSARINS-chem module mentioned above, has been completely re-designed and re-developed, to be user friendly and to facilitate the application and use of QSAR models. It does not require any license and can be freely downloaded at http://dunant.dista.uninsubria.it/qsar/.

A new simple graphic user interface (GUI) guides scientists and regulators step-by step in the application of more than 20 QSAR models to guarantee a straightforward and transparent procedure. The reliability of predictions can be analyzed using numerical and graphical outputs that highlight criticalities and allow for the investigation of the applicability domain of the models.

The QSARs included in the new software predict the potential hazard related to environmental fate, metabolism and toxicity as well as the potential Persistent, Bioaccumulative and Toxic (PBT) behavior of traditional and emerging contaminants (e.g., pharmaceuticals and personal care products [PCPs]).

Furthermore, the new QSARINS-Chem standalone version includes an updated version of the full database of molecular structures which was provided within the QSARINS-Chem module,[17] with new datasets added after 2014.

QSARINS-Chem standalone version is a useful tool dedicated to general users, not necessarily QSAR developers or QSAR experts, to generate screening-level information to support weight of evidence analysis in the context of chemicals risk assessment.

2 | METHODS

2.1 | Software design

To be platform independent, QSARINS-Chem standalone version was developed using the Java™ language. JavaFX™ was used for GUI layout and the 3D molecular drawing. To keep the software requirements to a minimum, the database of molecular structures is handled directly by the software, therefore no additional database server installation is needed. The models' database and the database of the molecular structures are organized and handled following the same scheme used for QSARINS-Chem module (2014), which is extensively explained in our previous publication.[17]

2.2 | Software overview

The new software is organized in two main modules: "Models" and "Database". The "Models" module allows the user to apply the models developed using QSARINS.[18] The "Database" module allows the user to browse the molecules in the database. These modules are accessible by two selectable tabs described below. In addition, we want to highlight that supporting documents called "Quick Start" and "How to" are available in the info section of the software, to easily guide users step by step through the application of the models and other functionalities.

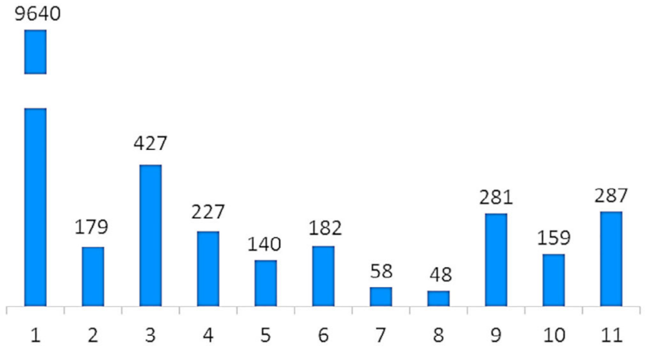

FIGURE 2 Vertical blue bars indicate the number of records included in each class of chemicals commented in the text (11 classes listed in Table 1)

2.2.1 | Models tab

The "Models" tab contains sub-tabs organized as a workflow for model's predictions. Once selected the following sub-tabs are shown:

1. "Models selection" tab. This tab allows for the selection of the model, using a drop-down list. For every selected model this tab shows the summary, the corresponding formula, the relevant statistics and validation criteria.
2. "Training descriptors" and "Training endpoint" tabs. Once a model is selected, the descriptors and the endpoint data of the model's training set are available for consultation in the "Training descriptors", and "Training endpoint" tabs. The last tab contains, in addition to the experimental values of the endpoint, the corresponding performances of the models (estimated values of the endpoint, the HAT value, normal and standardized residuals, also for cross validated values).
3. "User descriptors" tab. This tab allows the user to enter the descriptors values and optionally the experimental endpoints. Apart from manual editing, compound's descriptors can be calculated by an automatic query. In this case the open-source software PaDEL-Descriptor is automatically called and configured by QSARINS-Chem and will automatically calculate the required descriptors.
4. Additional tabs. Once data is entered and accepted, additional tabs are activated allowing the analysis of the model's predictions. The "Predictions" tab shows the predicted endpoint values (estimated endpoint) and relevant statistics. The "Graph" tab shows simultaneously (for an easier comparison) the graphs relevant to evaluate the model's performances, that is, experimental versus estimated endpoints, the corresponding residuals, the Williams plot and the "Insubria graph"[17,18], to examine the applicability domain of the models.

2.2.2 | Database tab

The "Database" tab shows the compound's relevant information, that is, name, CAS, SMILES, endpoint value and the related references, in a tabular form. For user's convenience, datasets can be viewed all together or dataset by dataset. When a single dataset is selected, it is possible to copy the SMILES of the compounds from the main page of the database or to export the compound's 3D structural files[17,21] where available. This tab also allows filtering the database by specific queries (name, molecular formula, CAS, and SMILES).

3 | DISCUSSION

3.1 | Structural database

A database consisting of 60 individual datasets for multiple endpoints and different chemical classes is included in the new software. This database is organized following the same structure used in the

TABLE 2 List of the QSAR models included in QSARINS-Chem standalone version

Category	Model	N of compounds in training set
Physico-chemical properties	1. Soil organic carbon-water partition coefficient (K_{OC})[17,22]	643
Global indexes	1. Global half-life index (GHLI)[17,23]	250
	2. Insubria PBT index[11,17]	180
Aquatic toxicity	1. Fish acute toxicity (*P. promelas*)[17,33]	449
Aquatic toxicity of personal care products (PCPs)	1. PCP freshwater Algae growth inhibition[7]	20
	2. PCP *Daphnia sp.* acute toxicity[7]	72
	3. PCP fish acute toxicity Model 1 (logP based)[7]	67
	4. PCP fish acute toxicity Model 2[7]	67
	5. PCP aquatic toxicity index (ATI)[7]	484
Aquatic toxicity of pharmaceuticals	1. Pharmaceutical freshwater Algae growth inhibition[8]	45
	2. Pharmaceutical *Daphnia* sp. acute toxicity[8]	125
	3. Pharmaceutical fish acute toxicity (*O. mykiss*)[8]	55
	4. Pharmaceutical fish acute toxicity (*P. promelas*)[8]	62
	5. Pharmaceutical aquatic toxicity index (ATI)[8]	706
Metabolic transformation in fish	1. Fish biotransformation Model 1[5]	632
	2. Fish biotransformation Model 2[5]	632
	3. Fish biotransformation Model 3[5]	632
Metabolic transformation in human	1. Human biotransformation Model 1[4]	1011
	2. Human biotransformation Model 2[4]	1015
	3. Human biotransformation Model 3[4]	935
	4. Human biotransformation Model 4[4]	940
	5. Human total elimination[4]	1105

QSARINS-Chem module (2014), described in our previous publication.[17] The database contains 3875 individual chemicals with 11,628 records (data points) grouped in nine main endpoint types (i.e., Physico-Chemical properties, Environmental Persistence, Bioconcentration Factor (BCF), Metabolic Transformation, Aquatic Toxicity, Terrestrial Toxicity, Biochemical Activity, Mutagenicity, and Endocrine Disruption). Table 1 and Figure 1 provide a brief overview of the structural database. We want to highlight that all these data were gathered from the literature or published databases (i.e., no new experimental data was measured in our laboratories). References listed in Table 1 refer to the original data (or databases) and/or to the curated datasets used to generate models included in the QSARINS software and in the QSARINS-chem Standalone version.

About half of the records are related to metabolic transformation data (more than 5000 records for about 1500 chemicals in fish and human).[28–32] Biotransformation is a fundamental component of bioaccumulation and information about biotransformation potential can reasonably be used to refine the bioaccumulation estimation.[49,50]

About one third of the remaining records are related to environmental persistence. These include data related to degradation half-life in sediment, soil, water and air, as well as data for atmospheric reactivity.

The third most relevant endpoint category is the aquatic toxicity. Datasets for this category include data for toxicity in different aquatic species (i.e., *Pseudokirchneriella subcapitata*, *Daphnia magna*, *Oncorinchus mykiss*, and *Pimephales promelas)* for PCPs, Pharmaceuticals, Benzotriazoles, Esters, and other heterogeneous organic chemicals.[7,8,37,38,43]

Physico-chemical properties are the fourth most relevant endpoint category. Particularly relevant is the dataset 1.1.1 in Table 1 for Soil organic carbon - water partition coefficient (K_{OC}), which counts 643 heterogeneous chemicals.[22] This dataset was used to generate a strongly externally validated QSAR model for the K_{OC} estimation of heterogeneous chemicals (see next section for further details).

Data for metabolic transformation in fish and human[28–32] have a relative high abundance. Finally, the rest of the records are relative to secondary small datasets for terrestrial toxicity,[44] ED,[34,40] and mutagenicity.[35,45]

The 3875 individual compounds and 11,628 records are divided in 11 classes according to the chemical class or to the classes of use (Figure 2) and cover a large and heterogeneous structural space. Class #1 is the largest and contains heterogeneous structures that were collected from large datasets. Classes from #2 to #11 are class-specific.

Class #1 is the only not class-specific group and covers 14 large datasets (among the largest in the database) for heterogeneous chemicals. It includes the Soil organic carbon - water partition coefficient (K_{OC}) dataset, the datasets related to persistence in multiple environmental media, the BCF datasets, the metabolic transformation in fish and human datasets, as well as datasets for the aquatic toxicity the endocrine disrupting properties.

Class #2 contains two datasets for the Mutagenicity Ames test TA100 of aromatic amines.

Class #3 is specific for (benzo)-triazoles and contains datasets for various physico-chemical properties and for aquatic toxicity.

Class #4 is specific for brominated flame retardants and contains the highest number of datasets: six datasets for physico-chemicals properties and seven datasets for different tests for endocrine disrupting activity.

Class #5 collects two datasets for aryl hydrocarbon receptor binding potency (pAH and pRB) of dioxin-like compounds.

Class #6 contains one dataset for physico-chemical properties and four datasets for aquatic toxicity of heterogeneous esters.

Class #7 is about fragrances and contains datasets for terrestrial toxicity in rat and some specific biochemical activity.

Class #8 contains record for mutagenicity of nitrated polycyclic aromatic hydrocarbons (PAH).

Class #9 is one of the largest and contains data for heterogeneous PFCs. This group gathers datasets describing physico-chemical properties and toxicities in rat and mice.

Classs #10 and #11 contain aquatic toxicity data on different organisms for PCPs and pharmaceuticals, respectively.

3.2 | QSAR models included in QSARINS-Chem Standalone version

The new QSARINS-Chem Standalone version includes 22 QSAR models, calculated using MLR by ordinary least squares, developed by the QSAR research group at University of Insubria in the last 20 years. These models, which are based on descriptors calculated by the PaDEL-Descriptor software, are listed in Table 2.

The models are grouped in seven different categories. A brief description of models included in each category is reported as follows. Readers can refer to specific literature reported in Table 2 and to the QMRF documentation provided in the software for further description.

Category 1 includes models for physico-chemical properties of heterogeneous compounds. The model for K_{OC} is characterized by good fitting, robustness and external predictivity and is based on a large structural domain which covers 663 compounds.

The second category includes QSARs generated for global indexes which encode for the tendency of chemicals to be persistent in the environment (i.e., persistent organic pollutants [POPs]) or to have a potential PBT behavior (i.e., chemicals which are persistent, bioaccumulative, and toxic). The Global Half-Life Index, namely the GHLI,[23] was derived by combining by principal component analysis (PCA) environmental half-lives in air, water, soil and sediments for 250 compounds, including POP-like chemicals. The PC1 score from the PCA is the GHLI index which was modeled by QSAR.[23,51] This model is helpful to screen new and existing chemicals and to identify potential POPs which should be substituted with less hazardous alternatives.

Similarly, the Insubria PBT Index[11,17,18] was calculated by PCA by combining data of environmental persistence, BCFs and acute toxicity measured in fish, for a set of 180 heterogeneous organic chemicals. A work by Sangion and Gramatica[10] has highlighted the utility of this QSAR model to screen the potential PBT behavior of pharmaceuticals.

Category #3 includes a model to estimate the aquatic toxicity of heterogeneous chemicals in fish based on the Duluth dataset of acute toxicity to *P. promelas*.[17,33,52,53]

Categories #4 and #5 include models to estimate the acute aquatic toxicity of PCPs and pharmaceuticals, respectively. These models estimate the toxicity of pharmaceuticals and PCPs in different test-standard organisms (i.e., *P. subcapitata*, *D. magna*, *O. mykiss*, and *P. promelas*) considered representative of different trophic levels in the aquatic environment. These models are intended to rank and prioritize PCPs and pharmaceuticals potentially toxic for aquatic environment in multiple species on the basis of the molecular structure; for instance, they can be used in the refinement of the toxicity assessment of a PBT screening. Models 4.5 and 5.5 (i.e., PCP aquatic toxicity index and pharmaceutical aquatic toxicity index) are particularly relevant because they are PCA-derived toxicity indexes specific for PCPs and pharmaceuticals; they combine toxicity information for alga, Daphnia and fish in a single index for the aquatic environment[7,10].

Finally, categories #6 and #7 include models for the biotransformation potential in fish and in human respectively. In particular, category #6 includes three models to predict *in vivo* whole-body biotransformation half-lives in fish. These models were developed on multiple training/prediction sets generated from a dataset composed of 632 compounds. The three models are based on different theoretical molecular descriptors and therefore have different structural applicability domains. We suggest using these models in a consensus approach, that is, by averaging predictions calculated by the three models.

Category #7 includes a model to estimate the human whole-body total elimination half-life and four models to predict the human whole-body primary biotransformation half-life. Model 7.1, 7.2, 7.3, 7.4 in Table 2 were developed on different training sets for the whole body biotransformation potential in human derived from different parametrizations of a 1-CoTK model.[32] These models can be applied

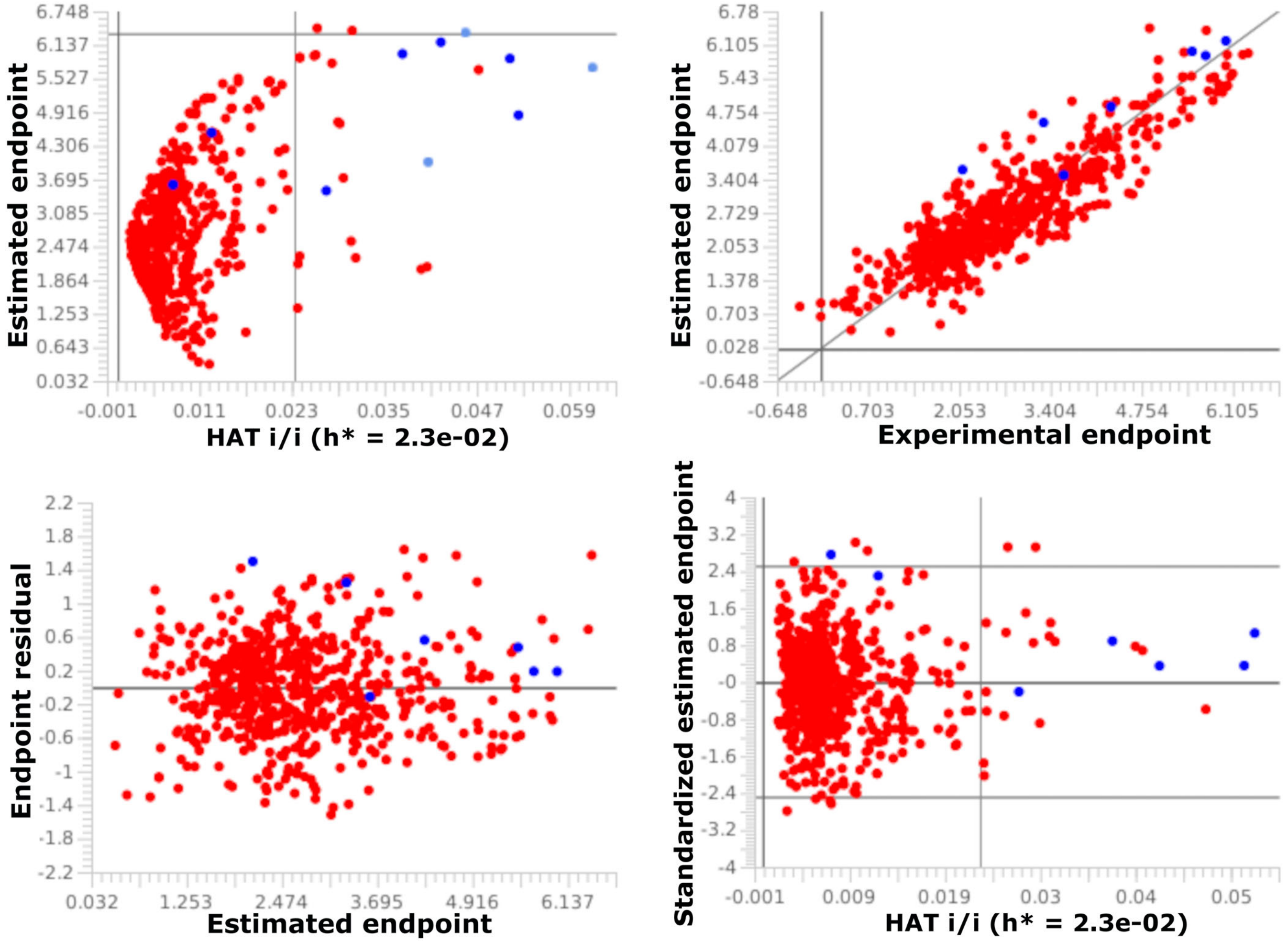

FIGURE 3 QSARINS-Chem diagnostic plots. Upper left: Plot of diagonal values from the HAT matrix versus estimated values of the endpoint ("Insubria graph"); upper right: Plot of experimental versus estimated values of the endpoint; lower left: Plot of estimated values of the endpointversus residuals; lower right: Plot of diagonal values from the HAT matrix versus standardized residuals ("Williams plot"). Colors: Red dots = training set; dark blue dots = user set, when the experimental value of the endpoint is provided by the user; light blue dots = user set, when the experimental value of the endpoint is not provided by the user

to estimate the biotransformation potential in fish and human and refine the bioaccumulation assessment of chemicals.

3.3 | Chemical profiling and diagnostics

The QSARINS–Chem standalone version drives the users through a transparent step-by-step procedure which goes from the selection of a QSAR model until the analysis of the reliability of predictions. Users can refer to specific literature reported for each model, to the QSARINS-Chem standalone manuals (i.e., "How to" and a "Quick-start guide"), and to the OECD Guidance on QSAR models development and validation[54] for further description of methods and parameters reported in the software.

The desired activity/property of one or more chemicals of interest for the user can be profiled singularly or in batch by applying the models listed in the software. In addition, the user can explore these predictions in comparison to chemicals used to train the selected QSAR by analyzing summary tables and graphs. Predictions are generated on the basis of molecular descriptors automatically calculated by the software. This is possible by uploading text files encoding for the molecular structure of the chemicals of interest (e.g., a list of SMILES with file extension .smi) in the PaDEL Descriptors Java application, which runs automatically through QSARINS-Chem Standalone version. Tables and graphs are exportable for personal use.

The procedure of selection and application of the models is summarized as follows:

Step 1: Model selection

The QSARINS-Chem standalone version opens as default on the model's selection page. A drop-down menu gives access to the desired model and to a description page, which summarizes all the main information associated to the selected model (i.e., model's description, equation, statistics, references). The QMRF document and the model file with experimental values and descriptors calculated for the training set chemicals (i.e., the .sdf file) can be exported from the model selection page.

The selection of a specific model activates three additional pages, that is, "Training descriptors", "Training endpoint", and "User descriptors". In particular, the first two pages are included for transparency to provide the information related to the input descriptors and the endpoint used to train the model as summary tables. Experimental and estimated values for the training set objects as well as residuals and leverage values (i.e., diagonal elements of the HAT matrix) are provided. Outliers for the response and influential objects are automatically highlighted in red in the "Training descriptors" and "Training endpoint" tables.

Step 2: Molecular descriptors calculation

The "User descriptors" page stores the values of the descriptors calculated for the molecules that are going to be profiled by the user. As mentioned above these descriptors can be automatically calculated within the QSARINS-Chem or entered manually.

Step 3: Predictions and diagnostics

The button "Apply model" available in the "User descriptors" page is used to run the model and generate predictions. This activates the "Predictions" page where are listed predicted values along with diagnostic parameters such as leverage values and residuals. These parameters are useful to address the reliability of the predictions. Values listed in the predictions page can be easily copied and pasted by the users for further analysis and storage.

Step 4: Graphical inspection

The "Graphs" page is automatically activated when predictions become available. Here the button "Calculate graphs" generates multiple graphs which allow to explore the position of the new prediction/predictions in the space of the original model. These plots are summarized in Figure 3 and can be exported by the user by simple copy-paste operation.

In particular, the here called "Insubria Graph"[17,18,37] (Figure 3—Upper left) plots HAT_{ii} (i.e., diagonal elements of the HAT matrix) versus estimated values of the endpoint, for the training set (red dots) and the user set (dark blue dots if the experimental value is provided by the user, otherwise light blue dots). This graph is useful to evaluate the inclusion of the user's chemicals in the applicability domain (structural space) of the model (i.e., compounds with $h_{i/i}$ values $\leq h^*$ are included in the applicability domain). In addition, predictions above or below the experimental range of the response (extrapolations) are also clearly identified.

Plot of the experimental versus estimated values of the endpoint (Figure 3—Upper right) provides visual information on the fitting of the model. If experimental values for the "User set" are feed in the software along with molecular descriptors they will appear in this graph (dark blue dots).

The plot of the residuals (Figure 3—Lower left) shows the estimated values of the endpoint versus residuals (where residuals = experimental values of the endpoint − estimated values of the endpoint). As was mentioned above if experimental values for the "User set" are provided, residuals calculated for the "User Set" will appear.

Finally, the Williams Plot (Figure 3—Lower right) shows the diagonal elements of the HAT matrix versus the standardized residuals. This graph provides combined information about the response and the structural domain and allows the identification of outliers.

4 | CONCLUSIONS

The new QSARINS-Chem Standalone version is a multiplatform, informative, and solid tool, which guides expert and non-expert users through a clear workflow for the QSAR-based profiling of existing and new chemicals taken singularly or in batch.

Compared to the original QSARINS-Chem module embedded in the QSARINS software, the new software has been fully redesigned and developed using the Java™ language. In addition, new functionalities are included such as summary tables, and diagnostic graphs that guarantee a straightforward and transparent use of the models. The over 20 QSAR models proposed in the software are particularly relevant to assess, through multiple endpoints, the PBT properties and the biotransformation of traditional and emerging contaminants. Furthermore, the analysis of the domain of the models is additionally supported by the structural database, which helps to address similarities among training chemicals and new compounds.

The reliability of predictions generated by the QSARINS-Chem standalone version can be analyzed using numerical and graphical outputs that highlight criticalities and allow for the investigation of the applicability domain of the models.

QSARINS-Chem standalone version provides valuable information to describe the behavior of chemicals at the screening level and to support hazard and risk assessment procedures for heterogeneous chemicals or specific structural and functional categories.

ACKNOWLEDGMENTS

We acknowledge Dr. Stefano Cassani for the revision of the chemical structures of the database (until 2015) and his contribution for the population of the QSARINS-Chem module with models based on PaDEL Descriptors. We acknowledge Dr. Jon Arnot for providing the Fish and the Human biotransformation data as well as Dr. James Armitage for helpful comments and for testing the first version of the software during BAT-Tool training session (Rome 2018). We gratefully acknowledge funding by CEFIC- LRI ECO44 that has supported part of the QSARINS-Chem development. We also acknowledge the PhD program in Chemical and Environmental Sciences (DiSCA, Università degli Studi dell'Insubria) for funding the scholarship of Linda Bertato.

ORCID

Ester Papa https://orcid.org/0000-0002-0041-556X

REFERENCES

[1] European Chemicals Agency (ECHA), Strategy to Promote Substitution to Safer Chemicals through Innovation, https://echa.europa.eu/documents/10162/13630/250118_substitution_strategy_en.pdf/bce91d57-9dfc-2a46-4afd-5998dbb88500 (accessed: April 12, 2021).

[2] United Nations Environment Programme (UNEP), Towards a Pollution-Free Planet Background Report, https://wedocs.unep.org/bitstream/handle/20.500.11822/21800/UNEA_towardspollution_long%20version_Web.pdf?sequence=1&isAllowed=y (accessed: April 12, 2021).

[3] D. R. Juberg, T. B. Knudsen, M. Sander, N. B. Beck, E. M. Faustman, D. L. Mendrick, J. R. Fowle, T. Hartung, R. R. Tice, E. Lemazurier, R. A. Becker, S. Compton Fitzpatrick, G. P. Daston, A. Harrill, R. N. Hines, D. A. Keller, J. C. Lipscomb, D. Watson, T. Bahadori, K. M. Crofton, *Toxicol. Sci.* **2017**, *155*, 22.

[4] E. Papa, A. Sangion, J. A. Arnot, P. Gramatica, *Food Chem. Toxicol.* **2018**, *112*, 535.

[5] E. Papa, L. van der Wal, J. A. Arnot, P. Gramatica, *Sci. Total Environ.* **2014**, *470*, 1040.

[6] K. Mansouri, N. Kleinstreuer, A. M. Abdelaziz, D. Alberga, V. M. Alves, P. L. Andersson, C. H. Andrade, F. Bai, I. Balabin, D. Ballabio, E. Benfenati, B. Bhhatarai, S. Boyer, J. Chen, V. Consonni, S. Farag, D. Fourches, A. T. García-Sosa, P. Gramatica, F. Grisoni, C. M. Grulke, H. Hong, D. Horvath, X. Hu, R. Huang, N. Jeliazkova, J. Li, X. Li, H. Liu, S. Manganelli, G. F. Mangiatordi, U. Maran, G. Marcou, T. Martin, E. Muratov, D. Nguyen, O. Nicolotti, N. G. Nikolov, U. Norinder, E. Papa, M. Petitjean, G. Piir, P. Pogodin, V. Poroikov, X. Qiao, A. M. Richard, A. Roncaglioni, P. Ruiz, C. Rupakheti, S. Sakkiah, A. Sangion, K. Schramm, C. Selvaraj, I. Shah, S. Sild, L. Sun, O. Taboureau, Y. Tang, I. V. Tetko, R. Todeschini, W. Tong, D. Trisciuzzi, A. Tropsha, G. Van Den Driessche, A. Varnek, Z. Wang, E. B. Wedebye, A. J. Williams, H. Xie, A. V. Zakharov, Z. Zheng, R. S. Judson, *Environ. Health Perspec.* **2020**, *128*, 27002.

[7] P. Gramatica, S. Cassani, A. Sangion, *Green Chem.* **2016**, *18*, 4393.

[8] A. Sangion, P. Gramatica, *Environ. Int.* **2016**, *95*, 131.

[9] A. Fernandez, A. Lombardo, R. Rallo, A. Roncaglioni, F. Giralt, E. Benfenati, *Environ. Int.* **2012**, *45*, 51.

[10] A. Sangion, P. Gramatica, *Environ. Res.* **2016**, *147*, 297.

[11] E. Papa, P. Gramatica, *Green Chem.* **2010**, *12*, 836.

[12] P. Gramatica, *IJQSPR* **2020**, *5*, 61.

[13] B. Bhhatarai, W. Teetz, T. Liu, T. Öberg, N. Jeliazkova, N. Kochev, O. Pukalov, I. V. Tetko, S. Kovarich, E. Papa, P. Gramatica, *Mol. Inf.* **2011**, *30*, 189.

[14] VEGA-QSAR. AI inside a platform for predictive toxicology, Proceedings of the workshop "Popularize Artificial Intelligence 2013", Turin, Italy, December 5th 2013, E. Benfenati, A. Manganaro, G. Gini, CEUR Workshop Proceedings Vol. *1107*, 21–28.

[15] Organization for the Economic Cooperation and Development (OECD). The OECD QSAR Toolbox. https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm (accessed: April 12, 2021).

[16] K. Mansouri, C. M. Grulke, R. S. Judson, A. J. Williams, *Aust. J. Chem.* **2018**, *10*, 10.

[17] P. Gramatica, S. Cassani, N. J. Chirico, *J. Comput. Chem.* **2014**, *35*, 1036.

[18] P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, *J. Comput. Chem.* **2013**, *34*, 2121.

[19] C. W. J. Yap, *Comput. Chem.* **2011**, *32*, 1466.

[20] Organization for the Economic Cooperation and Development (OECD). OECD Principles for the Validation, for Regulatory Purposes, of (Q)SAR Models. https://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf (accessed: April 12, **2021**).

[21] Chemistry Software, HyperChem, Molecular Modeling. http://www.hyper.com/ (accessed: April 12, 2021).

[22] P. Gramatica, E. Giani, E. Papa, *J Mol Graph Model.* **2007**, *25*, 755.

[23] P. Gramatica, E. Papa, *Environ. Sci. Technol.* **2007**, *41*, 2833.

[24] E. Papa, P. Gramatica, *SAR QSAR Environ. Res.* **2008**, *19*, 655.

[25] P. Gramatica, P. Pilutti, E. Papa, *QSAR Comb. Sci.* **2003**, *22*, 364.

[26] P. P. Roy, S. Kovarich, P. Gramatica, *J. Comput. Chem.* **2011**, *32*, 2386.

[27] P. Gramatica, E. Papa, *QSAR Comb. Sci.* **2005**, *24*, 953.

[28] J. A. Arnot, D. Mackay, M. Bonnell, *Environ. Toxicol. Chem.* **2008**, *27*, 341.

[29] J. A. Arnot, D. Mackay, T. F. Parkerton, M. Bonnell, *Environ. Toxicol. Chem.* **2008**, *27*, 2263.

[30] J. A. Arnot, W. Meylan, J. Tunkel, P. H. Howard, D. Mackay, M. Bonnell, R. S. Boethling, *Environ. Toxicol. Chem.* **2009**, *28*, 1168.

[31] T. N. Brown, J. A. Arnot, F. Wania, *Environ. Sci. Technol.* **2012**, *46*, 8253.

[32] J. A. Arnot, T. N. Brown, F. Wania, *Environ. Sci. Technol.* **2014**, *48*, 723.

[33] E. Papa, F. Villa, P. Gramatica, *J. Chem. Inf. Model.* **2005**, *45*, 1256.

[34] J. Li, P. Gramatica, *Mol. Diversity* **2010**, *14*, 687.

[35] P. Gramatica, V. Consonni, M. Pavan, *SAR QSAR Environ. Res.* **2003**, *14*, 237.

[36] B. Bhhatarai, P. Gramatica, *Water Res.* **2011**, *45*, 1463.

[37] P. Gramatica, S. Cassani, P. P. Roy, S. Kovarich, C. W. Yap, E. Papa, *Mol. Inf.* **2012**, *31*, 817.

[38] S. Cassani, S. Kovarich, E. Papa, P. P. Roy, L. van der Wal, P. Gramatica, *J. Hazard. Mater.* **2013**, *258*, 50.

[39] E. Papa, S. Kovarich, P. Gramatica, *QSAR Comb. Sci.* **2009**, *28*, 790.

[40] E. Papa, S. Kovarich, P. Gramatica, *Chem. Res. Toxicol.* **2010**, *23*, 946.

[41] R. Todeschini, P. Gramatica, *Quant. Struct.-Act. Relat.* **1997**, *16*, 120.

[42] P. Gramatica, F. Battaini, E. Papa, *Fresenius Environ. Bull.* **2004**, *13*, 1258.

[43] E. Papa, F. Battaini, P. Gramatica, *Chemosphere* **2005**, *58*, 559.

[44] E. Papa, M. Luini, P. Gramatica, *SAR QSAR Environ. Res.* **2009**, *20*, 767.

[45] P. Gramatica, P. Pilutti, E. Papa, *SAR QSAR Environ. Res.* **2007**, *18*, 169.

[46] B. Bhhatarai, P. Gramatica, *Environ. Sci. Technol.* **2011**, *45*, 8120.

[47] B. Bhhatarai, P. Gramatica, *Mol. Diversity* **2011**, *15*, 467.

[48] B. Bhhatarai, P. Gramatica, *Chem. Res. Toxicol.* **2010**, *23*, 528.

[49] M. S. McLachlan, G. Czub, M. MacLeod, J. A. Arnot, *Environ. Sci. Technol.* **2011**, *45*, 197.

[50] European Chemicals Agency (ECHA), Guidance on Information Requirements and Chemical Safety Assessment Chapter R.11: PBT/VPvB Assessment. https://echa.europa.eu/documents/10162/13632/information_requirements_r11_en.pdf (accessed: April 12, 2021).

[51] P. Gramatica, E. Papa, A. Sangion, *Environ. Sci.-Process Impacts* **2018**, *20*, 38.

[52] Fathead Minnow Dataset. https://archive.epa.gov/med/med_archive_03/web/html/fathead_minnow.html (accessed: April 12, 2021).

[53] C. L. Russom, S. P. Bradbury, S. J. Broderius, D. E. Hammermeister, R. A. Drummond, *Environ. Toxicol. Chem.* **1997**, *16*, 948.

[54] Organization for the Economic Cooperation and Development (OECD), Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. https://www.oecd-ilibrary.org/docserver/9789264085442-en.pdf?expires=1618216738&id=id&accname=guest&checksum=51796AE2A593B32FA9B9E8A122127151 (accessed: April 12, 2021).

How to cite this article: N. Chirico, A. Sangion, P. Gramatica, L. Bertato, I. Casartelli, E. Papa, *J Comput Chem* **2021**, *42*(20), 1452–1460. https://doi.org/10.1002/jcc.26551