QSARINS-Chem

Standalone Version

Quick Start Guide

Authors

Ester Papa (Group Leader)
Nicola Chirico (Software Developer)
Alessandro Sangion
Paola Gramatica (Senior Professor)

QSAR Research Unit in Environmental Chemistry and Ecotoxicology
Department of Theoretical and Applied Sciences (DiSTA)
University of Insubria, Varese, Italy
http://dunant.dista.uninsubria.it/qsar/

Information about the Standalone Version

Models

This standalone version of QSARINS-Chem contains QSAR models for the prediction of:

Category	Model
1. Physico-Chemical properties	1. Soil Organic Carbon-Water partition Coefficient (K_{OC})
2. Global Indexes	1. Global Half-Life Index 2. Insubria PBT Index
3. Aquatic Toxicity	1. Fish Acute Toxicity (*P.promelas*)
4. Aquatic Toxicity of Personal Care Products (PCPs)	1. PCP Freshwater Algae Growth Inhibition 2. PCP *Daphnia* sp. Acute Toxicity 3. PCP Fish Acute Toxicity Model 1 (logP based) 4. PCP Fish Acute Toxicity Model 2 5. PCP Aquatic Toxicity Index (ATI)
5. Aquatic Toxicity of Pharmaceuticals	1. Pharmaceutical Freshwater Algae Growth inhibition 2. Pharmaceuticals *Daphnia* sp. acute Toxicity 3. Pharmaceuticals fish Acute Toxicity (*O.mykiss*) 4. Pharmaceuticals fish Acute Toxicity (*P.promelas*) 5. Pharmaceuticals Aquatic Toxicity Index (ATI)
6. Metabolic transformation in fish	1. Fish Biotransformation Model 1 2. Fish Biotransformation Model 2 3. Fish Biotransformation Model 3
7. Metabolic transformation in human	1. Human biotransformation Model 1 2. Human biotransformation Model 2 3. Human biotransformation Model 3 4. Human biotransformation Model 4 5. Human Total elimination

Database

This standalone version of QSARINS-Chem contains the database published by Gramatica et al. 2014[1] which includes the following datasets:

Chemical Class	Endpoint-Type	Dataset Name
1. General	1. Physico-Chemical Properties	1. Soil Organic Carbon-Water partition Coefficient (K_{OC})
	2. Environmental Persistence	1. Sediment Half-Lives 2. Soil Half-Lives 3. Water Half-Lives 4. Air Half-Lives 5. NO_3 reactivity 6. O_3 reactivity 7. OH reactivity 8. Global Half-Life Index
	3. Bioconcentration Factor	1. BCF-Fernandez 2. BCF-Lu
	4. Metabolic Transformation	1. Fish Biotransformation
	5. Aquatic Toxicity	1. Fish acute toxicity (*P.promelas*)
	6. Endocrine Disruption	1. Estrogen Receptor Binding
2. Aromatic Amines	1. Mutagenicity	1. Aromatic Amines mutagenicity TA98 2. Aromatic Amines mutagenicity TA100
3. (Benzo)Triazoles	1. Physico-Chemical Properties	1. (B)TAZ Kow 2. (B)TAZ Solubility in Water 3. (B)TAZ Vapor Pressure 4. (B)TAZ Melting Point
	2. Aquatic Toxicity	1. (B)TAZ Algae acute toxicity (*P.subcapitata*) 2. (B)TAZ *Daphnia sp* acute toxicity 3. (B)TAZ Fish acute toxicity (*O.mykiss*)
4. Brominated Flame Retardants	1. Physico-Chemical Properties	1. BFR Kow 2. BFR Koa 3. BFR Vapor Pressure 4. BFR Solubility in Water 5. BFR Henry Law Constant 6. BFR Melting Point
	2. Endocrine Disruption	1. BFR-DR-Ag 2. BFR-ER-Ag 3. BFR-ERODind 4. BFR-PR-ant 5. BFR-SULT-REP 6. BFR-T4-REP 7. BFR Receptor Binding Affinity

Chemical Class	Endpoint-Type	Dataset Name
5. Dioxin Analogues	1. Biological Activity	1. Dioxin Analogues pAHH 2. Dioxin Analogues pRB
6. Esters	1. Physico-Chemical Properties	1. Esters Flash Point
	2. Aquatic Toxicity	1. Esters Algae acute toxicity 2. Esters *Daphnia sp* acute toxicity 3. Esters Fish acute toxicity (*P.promelas*) 4. Esters Aquatic Toxicity Index (EATIN)
7. Fragrances	1. Terrestrial Toxicity	1. Fragrances Oral toxicity (Rat)
	2. Biochemical activity	1. Fragrances Inhibition NADHox 2. Fragrances Mitochondrial memb pot
8. Nitrated polycyclic aromatic hydrocarbons	1. Mutagenicity	1. NitroPAH mutagenicity TA100
9. Perfluorinated Compounds	1. Physico-Chemical Properties	1. PFC Critical Micelle Concentration 2. PFC Solubility in Water 3. PFC Vapor Pressure
	2. Terrestrial Toxicity	1. PFC Oral toxicity (Rat) 2. PFC Oral toxicity (Mouse) 3. PFC Inhalation toxicity (Rat) 4. PFC Inhalation toxicity (Mouse)
10. Personal Care Products	1. Aquatic Toxicity	1. PCP Algae acute toxicity (*P.subcapitata*) 2. PCP *Daphnia sp* acute toxicity 3. PCP Fish acute toxicity (*P.promelas*)
11. Pharmaceuticals	1. Aquatic Toxicity	1. Pharm. Algae acute toxicity (*P.subcapitata*) 2. Pharm. *Daphnia sp* acute toxicity 3. Pharm. Fish acute toxicity (*O.mykiss*) 4. Pharm. Fish acute toxicity (*P.promelas*)

NOTE ABOUT DATA AND STRUCTURES: All the chemical structures have been designed and "energetically minimized" (AM1 method in HyperChem v.7.03) by the QSAR Research Unit in Environmental Chemistry and Ecotoxicology at the University of Insubria[1]. SMILES (not in the canonical form) reported in the database have been generated from the 3D structures using OpenBabel v 2.3.2. All the experimental data have been collected from literature. Information about original data, data collection and curation can be found in the publications cited in the database. No experimental data have been generated by QSAR Research Unit in Environmental Chemistry and Ecotoxicology at the University of Insubria (exceptions are the global indexes generated by Principal Component Analysis).

---

[1] "QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS", Gramatica P., Cassani S., Chirico N. Journal of Computational Chemistry 2014, 35, 1036-1044. DOI: 10.1002/jcc.23576

Acknowledgments

How to cite:

Please cite QSARINS-Chem standalone version in your publications as:

Papa Ester, Chirico Nicola, Sangion Alessandro, Gramatica Paola, QSARINS-Chem standalone version: a free platform-independent software for QSAR-based predictions of properties and activities of organic chemicals, submitted to J.Comput. Chem.

We would also appreciate citations to the website: http://dunant.dista.uninsubria.it/qsar/

Limitations of Liability and Disclaimer of Warranty

Overview

QSARINS-Chem allows the user to input his molecular structures and get estimations and Applicability Domain for a desired QSAR model.

Minimal information required to obtain prediction are:

-Chemical structures files (e.g. .smi, .mol) (see below to see how to set a .smi file)

Step 1: Select the model:

1) **Open** QSARINS-Chem.jar. The first time you use it, you need to accept the "Licence agreement" selecting "I agree".

2) The "Model Selection" page will open; here **you can select** the desired **QSAR model** from the blue drop-down menu. This page summarizes the main information of the selected model (model's description, equation and statistics).

3) The following two tabs ("Training descriptors" and "Training endpoint") give information about the Training set of the model:

> The "Training descriptors" tab reports the values of the descriptors of the Training set of the selected QSAR model.

> The "Training endpoint" tab reports Experimental and Estimated endpoint for the Training set objects as well as Residuals and Leverage values (i.e. diagonal elements of the Hat matrix). Outliers for the response and influential objects are highlighted in red.

Step 2: Enter molecular descriptors and predict your data

1) **Select** "User descriptors" tab to initialize your chemicals and calculate molecular descriptors.

2) **Right Click** on the first empty line and **select** "Import and calculate descriptors" from the drop-down menu.

3) **Select** the directory containing the molecules' structural files. The software **PadelDescriptor** will calculate and load molecular descriptors for your chemicals.

Optional: If available **you can enter the experimental value of your response (in the same units of the selected QSAR models). **Type** your value/values in the "User Response" column and **press** "Enter".*

Warning: The column "Status" will report the presence of issues in your data with a red warning (i.e. descriptors and/or User response out of range of the Training set or **missing descriptors**). **QSARINS-Chem cannot process molecules with missing descriptors; manually enter the value or delete the chemical (right click and press delete).

4) **Press** "Apply Model" to run the model and generate predictions.

Step 3: Get Estimated Values and Applicability Domain

1) The "Predictions" tab will be activated; this tab reports Estimated values and Leverages (i.e. diagonal elements of the Hat matrix) for the user set. Residuals will be calculated only if user experimental values are provided. The column "Status" gives indication about possible issues related to the Applicability Domain (i.e. descriptors and/or User response out of range of the training set or Leverages higher than the cut-off value h*).

2) **Select** the chemicals of your interest and **right click to copy** the predictions. **Paste** in "Excel" or in other format to save the estimations.

Step 4: Graphical inspection

1) In the "Graphs" tab **press** "Calculate Graphs" button to generate graphs:

-Insubria Graph: plot of Leverages (i.e. diagonal elements of the Hat matrix) vs Estimated values or training set (red) and User set (blue). Here you **can evaluate** the inclusion of the User's chemicals in the Applicability Domain (structural space) of the model. Compounds with $h_{i/i}$ values <= h* are included in the Applicability Domain. In addition you **can evaluate** chemicals estimated above or below the estimated response range of the training set.

-Experimental vs Estimated values: this plot provides visual information on the fitting of the model. If Experimental values for the "User set" are provided they will appear in this graph.

-Residual plot: this graph plots the Experimental Response vs the Residuals (Experimental Response – Estimated Response). If Experimental values for the "User set" are provided, the "User Set" will appear in blue.

- William's Plot: this graph plots the Experimental Responses vs the Standardized Residuals. Here you **can evaluate** outliers for the response and influential objects.

2) If interested, you **can copy and paste or save** any of these graphs.

How to set up a SMILES structural file (.smi)

PadelDescriptor software can generate descriptors from SMILES placed in a **.smi** structural file. This is a "tab delimited" text file containing the **SMILES structures in the first column** and the Identifiers (optional) in the second column, **no header**. The extension of file must be **.smi**

To generate this file in "Excel" (or similar software):

1) **Open** an empty document

2) **Paste** your SMILES in the first column

3) If present, **paste** the identifier (ID or CAS or NAME) in the second column

4) If present, **delete** the header

5) **Save** as tab delimited text file [*TEXT (tab delimited)(*.txt)*]

6) Manually **change** the extension from **.txt** to **.smi**

(*To see file extension on*
***Windows 7: Open Windows Explorer** and click the **Organize** button towards the top left. Choose **Folder and search options** from the menu. Click the **View** tab in the window that opens, then scroll down and untick the box next to 'Hide file extensions for known file types'*
***Windows 8/10**: open a **File Explorer** window (the new name for Windows Explorer) and click the **View** tab.*
***Mac**: Click on the **Finder** menu and select **Preferences**. Select **Advanced** button and put a check mark in the checkbox labeled **Show all filename extension**)*

7) **Place** the file in an empty folder to use in **Step 2 point 3**