Re-Thinking Experimentation

QSAR Model Reporting Format (QMRF) for iSafeRat® (High Accuracy)-QSAR for acute ecotoxicological endpoints – Version 2.6

1 QSAR identifier

1.1 QSAR identifier (title)

iSafeRat® algErC50
iSafeRat® daphEC50
iSafeRat® fishLC50

1.2 Other related models

None

1.3 Software coding the model

The ecotoxicological models are implemented to the commercial software iSafeRat® Desktop.
The version of the ecotoxicological models presented in this QMRF are present in the version 5.2.3 and later versions of iSafeRat® Desktop.
Additional documentation: KREATiS (2023) – iSafeRat® Desktop v5.2.3 User Guide

2 General information

2.1 Abstract

This QMRF is related to all the acute aquatic ecotoxicological models implemented to the iSafeRat® desktop commercial software. The QMRF deals with three families of model: iSafeRat® algErC50, iSafeRat® daphEC50 and iSafeRat® fishLC50. Each family of model contains a specific number of individual models (all are simple linear regressions). The models were developed to predict acute ecotoxicity values for fish, daphnids and algae that would be expected in a laboratory experiment following the OECD Guideline 203 and 236 (OECD, 2019, 2013), 202 (OECD, 2004) and 201 (OECD, 2011), respectively. The data present in the models were obtained from public sources in most cases and were checked internally, manually one by one to be sure that the experimental data are valid according to the OECD guidelines and that the experimental settings and parameters were adapted to the properties of the tested compounds. Those that did not meet the strict criteria were not considered fit for purpose regardless of the external validation scores given to them. Only organic compounds were considered in the models. Each aquatic toxicity model developed by KREATiS experts are implemented in the software iSafeRat® as simple linear-relationships between the Subcooled Liquid Water-Solubility (SLWS) (with the exception of the quaternary ammonium compounds where the independent variable is the octanol-water partition-coefficient) and the ecotoxicity. Moreover, all these models were based on a mechanistic interaction between the chemical and the organisms (MechoA, Bauer et al., 2018). Therefore, the descriptors in one regression can contain more than one functional chemical group if both have the same MechoA and are recognised as acting similarly based on scientific knowledge. The iSafeRat® algErC50, iSafeRat® daphEC50 and iSafeRat® fishLC50 contains 9, 10 and 11 independent models, respectively (one per MechoA). The statistics assessing the goodness-of-fit, the robustness and the predictivity of the models were calculated according to the OECD (2007) document and are described below for transparency.

2.2 Date of QMRF

16 June 2025

2.3 Date of QMRF update(s)

Table 1: Dates of QMRF updates

Date	QMRF update identifier
16 June 2025	KTS/QMRF/ACU/05
20 January 2025	KTS/QMRF/ACU/04
30 April 2024	KTS/QMRF/ACU/03
20 November 2023	KTS/QMRF/ACU/02
04 April 2023	KTS/QMRF/ACU/01

2.4 QMRF update(s)

Table 2: Contents of QMRF updates

QMRF update identifier	Content
KTS/QMRF/ACU/05	• Update of the following ecotoxicological model: MechoA 1.2 algae: addition of a dianiline compound in the training set.
KTS/QMRF/ACU/04	• Update of the following ecotoxicological models: MechoA 1.1 fish: addition of 2 amide, 1 lactam and 2 fluoride compounds in the training set and of 1 amide, 1 lactam and 1 fluoride compounds in the validation set. MechoA 1.1 daphnids : addition of 1 amide, 1 lactam and 2 fluoride compounds in the training set. MechoA 1.1 algae : addition of 2 amide, 2 lactam and 2 fluoride compounds in the training set of the model. Addition of 1 amide and 2 fluoride compounds in the validation set. • Update of structural fragment space of each model based on the update list of SMARTS used in the iSafeRat® models.
KTS/QMRF/ACU/03	• Update of QMRF format according to the QAF template (version 2.4). • Update of the information provided in sections 6 and 7. • Update of the following ecotoxicological models: MechoA 1.1 fish, daphnids and algae (redundant compounds removal); MechoA 1.2 algae: Due to a limited number of reliable data, training and validation sets were merged to obtain a more robust model; MechoA 2.1 monoester fish: addition of 3 phosphate compounds in the training set and exclusion of a statistical outlier (based on residuals); MechoA 2.1 monoester daphnids: addition of 3 phosphate compounds and transfer of several compounds of the training set to the validation set to render both dataset more balanced in term of number of compound; MechoA 2.1 ester algae: addition of 4 phosphate compounds in the training set and transfer of several compounds of the training set to the validation set to render both dataset more balanced in term of number of compound; MechoA 3.2 fish and algae: internal correction; MechoA 2.1 polyester daphnids, MechoA 5.2 daphnids and algae, MechoA 1.2&5.2 daphnids, MechoA an1.2&4.3 polar anilines: Due to a limited number of reliable data, training and validation sets were merged to obtain a more robust model.
KTS/QMRF/ACU/02	Response domains information update.
KTS/QMRF/ACU/01	Most of the ecotoxicological model presented in this new QMRF were previously presented in the previous QMRF entitled (KREATiS QMRF ID: KTS/QMRF/HOL/09). This new QMRF includes only acute ecotoxicological models. Three models were updated and 5 models were newly developed. • Updated models: - algal, daphnid and fish acute toxicity models for MechoA 3.1 epoxide integration to models) - algal acute toxicity models for MechoA 1.2&5.2 (amines) • Newly developed models:

	- algal, daphnid and fish acute toxicity models for MechoA 3.2 (α,β-unsaturated esters and aldehydes); - fish acute toxicity models (one for simple anilines and one for polar anilines) for MechoA an4.3&1.2.
KTS/QMRF/HOL/09	Originally, the acute ecotoxicological QSAR models were included in the so-called holistic QMRF grouping QSARs for several interconnected endpoints. The first version of the holistic QMRF, released on 27 February 2014, already included the acute ecotoxicological QSAR. KTS/QMRF/HOL/09 was the last version of the holistic QMRF, after several updates to all models.

2.5 Model developer(s) and contact details

KREATiS SAS
23, rue du Creuzat
38080 L'ISLE D'ABEAU
France
Tel: +33 (0)4 28 19 01 06
Email: contact@kreatis.eu
Website: www.kreatis.eu

2.6 Date of model development and/or publication

The results presented in this QMRF refer to the current version of the following models:
- iSafeRat® fishLC50 – v2.1
- iSafeRat® daphEC50 – v2.1
- iSafeRat® algErC50 – v2.2

All models were validated on 16th June 2025.

2.7 Reference(s) to main scientific papers and/or software package

The ecotoxicological models are part of iSafeRat® – in Silico Algorithms For Environmental Risk And Toxicity, available through the following software tools:

- iSafeRat® Desktop version 5.2.3 and later versions. iSafeRat® Desktop version 5.2.3 for High-Accuracy QSAR predictions by KREATiS SAS. https://isaferat.kreatis.eu/

2.8 Availability of information about the model

The models are proprietary but limited information has been made publicly available for the validation/test set. Any queries concerning the model or its validity should be addressed at contact@kreatis.eu. Furthermore, KREATiS undertakes to provide supplementary information to sponsors or regulatory authorities upon request to demonstrate compliance of our QSARs with good practice.

2.9 Availability of another QMRF for exactly the same model

None

3 Defining the endpoint - OECD Principle 1

3.1 Species

The species for which relevant studies were included in the model's training set for the three ecotoxicological endpoints were:

- Short-term toxicity to fish: *Danio rerio, Lepomis macrochirus, Pimephales promelas, Oncorhynchus mykiss Oryzias latipes, Poecilia reticulata, Leuciscus idus, Cyprinus carpio*
- Short-term toxicity to invertebrates: *Daphnia magna, Daphnia pulex*
- Toxicity to algae: *Desmodesmus subspicatus (synonyme: Scenedesmus subspicatus), Scenedesmus quadricauda, Chlorella vulgaris, Raphidocelis subcapitata (synonyme: Pseudokirchneriella subcapitata)*

3.2 Endpoint

The models can accurately predict the acute aquatic toxicity to:

- Algae, based on the inhibition of the exponential growth rate that would be expected in a laboratory experiment following the OECD Guideline 201.
- Daphnid, based on the immobility of the individuals that would be expected in a laboratory experiment following the OECD Guideline 202.
- Fish, based on the mortality of the individuals that would be expected in a laboratory experiment following the OECD Guideline 203 and 236.

3.3 Comment on endpoint

The endpoints represent i) the 50% Effect Concentration for the fish mortality (LC50), ii) the daphnids immobilization (EC50) and iii) the algal growth inhibition (ErC50).

3.4 Endpoint units

The predicted endpoint value (LC50 or EC50 or ErC50) is first derived from the model in log 10 (mol/L), then converted in mol/L and finally given in **mg/L**.

3.5 Dependent variable

Table 3 provides an overview of the dependent variables of all the endpoints involved (and additionally, test durations are provided for all ecotoxicological endpoints).

Table 3: Endpoints and model dependent variables

Endpoint	Dependent Variable
Acute toxicity to fish *(lethality)*	96h-LC50 value in log10 (mol/L) (*Median Lethal concentration*)
Short-term toxicity to Daphnia *(immobilisation)*	48h-EC50 value in log10 (mol/L) (*Median effective concentration*)
Short-term toxicity to algae *(inhibition of the exponential growth rate)*	72h-ErC50 value in log10 (mol/L) (*Median inhibition concentration*)

3.6 Experimental protocol

- **Dependent variable (fish mortality or daphnids immobility or algal growth rate)**

The ecotoxicity values were obtained from studies which were carried out using the method recommended in the OECD guideline 203 and 236 for fish (96h-LC50 values), 202 for aquatic invertebrates (48h-EC50 values), and 201 for algae (72h-ErC50 values). The respect of the validity criteria of the experimental studies (e.g. organisms stability in control over-time; physicochemical condition stability over-time; satisfying concentration recovery if nominal based ecotoxicological values) was investigated. Also, the adequacy between experimental settings and the properties of the tested compounds was checked.

- **Independent variable (descriptor: subcooled liquid water solubility or octanol-water partition coefficient)**

For all models except one, the descriptor of the model is the subcooled liquid water solubility (SLWS, calculation explained in section 4.3) of the compounds. The data used for water solubility values come from studies which have been carried out using the method recommended in the OECD guideline 105, *i.e.,* shake-flask (OECD, 1995a). An adaptation (slow-stir method) is recommended for liquids with an expected solubility below 10 mg/L (European Chemicals Agency, 2016; Letinski et al., 2002), and data with this method have been used as often as possible. The aim of this method is to prevent the formation of emulsions that can occur with the shake flask method and lead to a solubility overestimation. In rare cases data has been taken from studies performed using a second method described in OECD guideline 105: the column elution method which is recommended for solids (not liquids) with water solubility below 10 mg/L.

To calculate the subcooled liquid water solubility value, the state liquid vs. solid of the compounds at 25°C is considered. The melting point values come from studies which have been carried out using one of the methods

recommended in the OECD guideline 102, i.e., capillary, Kofler hot bar, melt microscope, differential thermal analysis, differential scanning calorimetry, freezing temperature or pour point method.

For the models for quaternary ammonium, the descriptor is the octanol-water partition coefficient. The data used for log KOW values were obtained from studies which were carried out using the method recommended in the OECD guideline 107 or 123, i.e., shake flask or slow-stirring method for liquids with expected solubility below 10 mg/L (OECD, 1995b, 2022a). Data coming from OECD guideline 117, i.e., HPLC method (OECD, 2022b) have been avoided because they are considered screening data and generally not accurate enough due to the high dependency to the calibration and the reference compound unless more appropriate reference compounds have been used for calibration.

3.7 Endpoint data quality and variability

One of the essentials to derive a HA-QSAR model is to perform a strict validation of the study results included within the training set. This validation was carried out using expert judgement by the dedicated KREATiS model development team. If for any reason the quality of the study results was compromised (for instance, due to unacceptable experimental conditions or issues with laboratory protocol), their corresponding results were withdrawn from the training set and the reason for their removal labelled in the internal database. The training set data comprised of quality results derived from multiple laboratories and as a result, inter-laboratory differences may be expected. For such cases, the results were not simply averaged but the validity of each result was then justified on a case by case basis. Cases with large differences in validated values for the same compound were treated with best possible modelling practice (for instance, potential outlier detection, data verification from available literature resources).

4 Defining the algorithm - OECD Principle 2

4.1 Type of model

Each model is based on a Simple Linear Regression.

4.2 Explicit algorithm

In each model, this approach correlates the Subcooled Liquid Water Solubility (SLWS) of a compound to the acute toxicity. For compounds with a non-polar narcotic mechanism of action (MechoA 1.1) according to the MechoA Scheme (Bauer et al., 2018), the predicted toxicity is based on a thorough understanding of the thermodynamic relationship between the activity and toxicity that is well discussed in the literature (Mackay et al., 2009; Thomas et al., 2015). For other MechoAs, a similar relationship is considered between SLWS values and ecotoxicological effects because SLWS is still a relevant descriptor. Indeed, the regressions will heavily rely on the MechoA and therefore, depending on different series of MechoA the obtained slopes and intercepts may vary significantly. Depending on the mechanism of action and the structural profile of the input chemical, it will be allocated to one of the local QSAR acute aquatic toxicity models listed in section 5.1.e.

The model algorithm is an affine equation of the type:

$$\log EC50 \; or \; LC50 = a \times \log SLWS + b$$

Where,
$EC50$: Effective Concentration leading to 50% of effect (mol.L^{-1}).
$LC50$: Effective Concentration leading to 50% of mortality (mol.L^{-1}).
$SLWS$: Subcooled Liquid Water Solubility (mol.L^{-1})

4.3 Descriptors in the model

The ecotoxicological models were based on the clear understanding of the thermodynamic relationship between ecotoxic effects and SLWS as the only descriptor. The selection of SLWS as a suitable descriptor for this model was based on the clear understanding of the thermodynamic relationship between the ecotoxicological effects and the chemical activity as proposed by (Mackay et al., 2009).

Therefore, the experimental water solubility values are corrected to take into account the SLWS which is an expression of the chemical activity of a pure compound within a given medium. The choice of the SLWS is

justified in order to compare liquids and solids within the same dataset. Indeed, the solids have an entropy of fusion different to liquids as demonstrated by (Yalkowsky, 1979). Consequently, the water solubility of the solids is corrected using their melting point according to the following equation:

$$\mathrm{SLWS}_{solid} = \frac{WS_{solid}}{e^{(6.79\left(1-\frac{MP}{298}\right))}}\ (mol.L^{-1})$$

SLWS_{solid} : Subcooled Liquid Water Solubility (SLWS) of a given compound with solid state at 25°C
WS_{solid}: Water Solubility (SLWS) of a given compound with solid state at 25°C (as measured).
MP: Melting Point of the given compound (K)

For liquids, the SLWS is directly equal to the value of the water solubility as measured.

4.4 Descriptor selection

The mechanistic understanding of the thermodynamic principles was the driving factor for descriptor selection. No other variable selection approaches were implemented. For each iSafeRat® ecotoxicological model, the most important aspect is to achieve the highest accuracy possible without including inexplicable molecular descriptors and any other mechanistically unjustified complexities.

4.5 Algorithm and descriptor generation

No specific variable selection method was applied. No statistical tool or packages were used to generate a pool of molecular descriptors. Only validated experimental study results were used (for independent and dependent variables).

4.6 Software name and version for descriptor generation

The descriptor (independent variable) and endpoint values (dependent variable) were either experimentally derived and retrieved from various literature resources including some publicly disseminated databases as well as some confidential data available within KREATiS internal database, or, for the water solubility values, some were predicted using iSafeRat® WatSol model, ensured to be within its applicability domain.

4.7 Chemicals/Descriptors ratio

The **tables 4.1, 4.2 and 4.3** provide with the Chemicals/Descriptors ratio for the QSAR models. Since the ratios were quite higher (equal to the number of training set compounds as the descriptor was equal to 1 in all the cases), it indicates that the models were not over fitted with a large number of descriptors.

Table 4.1: Chemicals/Descriptor ratios for different iSafeRat® fishLC50 models

MechoA	Fish acute ecotoxicity models	Chem/Desc. ratio
Non-polar narcosis (MechoA 1.1)	iSafeRat® fishLC50	71
Polar-narcosis (MechoA 1.2)	iSafeRat® fishLC50	16
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	iSafeRat® fishLC50	35
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	iSafeRat® fishLC50	18
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® fishLC50	15
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® fishLC50	12
Redox cycling (MechoA 4.4)	iSafeRat® fishLC50	6
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® fishLC50	11

Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® fishLC50	14
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, simple anilines)	iSafeRat® fishLC50	12
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, polar anilines)	iSafeRat® fishLC50	14

Table 4.2: Chemicals/Descriptor ratios for different iSafeRat® daphEC50 models

MechoA	Daphnids acute ecotoxicity models	Chem/Desc. ratio
Non-polar narcosis (MechoA 1.1)	iSafeRat® daphEC50	61
Polar-narcosis (MechoA 1.2)	iSafeRat® daphEC50	19
Narcosis of permanently cationic molecules (MechoA 1.3)	iSafeRat® daphEC50	7
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	iSafeRat® daphEC50	46
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	iSafeRat® daphEC50	13
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® daphEC50	30
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® daphEC50	16
Redox cycling (MechoA 4.4)	iSafeRat® daphEC50	7
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® daphEC50	15
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® daphEC50	16

Table 4.3: Chemicals/Descriptor ratios for different iSafeRat® algErC50 models

MechoA	Algal acute ecotoxicity models	Chem/Desc. ratio
Non-polar narcosis (MechoA 1.1)	iSafeRat® algErC50	45
Polar-narcosis (MechoA 1.2)	iSafeRat® algErC50	14
Narcosis of permanently cationic molecules (MechoA 1.3)	iSafeRat® algErC50	4
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono and poly-ester)	iSafeRat® algErC50	49
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® algErC50	15
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® algErC50	13
Redox cycling (MechoA 4.4)	iSafeRat® algErC50	4
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® algErC50	12

Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® algErC50	16

5 Defining the applicability domain - OECD Principle 3

5.1 Description of the applicability domain of the model

a) Fixed or probabilistic boundaries

The following domains are fixed boundaries based on training set data.

b) Response domain

The model is limited by its response with a minimum and a maximum value of ecotoxicity (ErC50 or EC50 or LC50 depending on the trophic level) defining a zone of interpolation. Beyond these extreme values, the prediction is an extrapolation. The quantified limits are given in the **Tables 5.1, 5.2 and 5.3:**

Table 5.1: Response domain for the different iSafeRat® fishLC50 ecotoxicological local models

MechoA	Fish acute ecotoxicity models	Response domain (LC50 in log10 (mol/L)) MIN	MAX
Non-polar narcosis (MechoA 1.1)	iSafeRat® fishLC50 - Zone 1	-5.6529	-1.1432
	iSafeRat® fishLC50 - Zone 2	-6.3840	-4.9380
	iSafeRat® fishLC50 - Zone 3	-8	0
Polar-narcosis (MechoA 1.2)	iSafeRat® fishLC50	-5.6156	-3.5774
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	iSafeRat® fishLC50	-5.4132	-2.2811
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	iSafeRat® fishLC50	-5.4903	-3.2908
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® fishLC50	-5.1385	-3.0480
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® fishLC50	-5.1155	-4.0172
Redox cycling (MechoA 4.4)	iSafeRat® fishLC50	-6.6467	-3.3246
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® fishLC50	-3.7563	-2.5599
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® fishLC50	-6.2183	-2.4360
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, simple anilines)	iSafeRat® fishLC50	-4.7599	-2.1270
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, polar anilines)	iSafeRat® fishLC50	-5.6443	-3.5089

Table 5.2: Response domain for the different iSafeRat® DaphEC50 ecotoxicological local models

MechoA	Daphnids acute ecotoxicity models	Response domain (EC50 in log10 (mol/L)) MIN	MAX
Non-polar narcosis (MechoA 1.1)	iSafeRat® daphEC50 – Zone 1	-5.5100	-1.5251
	iSafeRat® daphEC50 – Zone 2	-5.7826	-5.2851
	iSafeRat® daphEC50 – Zone 3	-8	0
Polar-narcosis (MechoA 1.2)	iSafeRat® daphEC50	-5.6619	-3.8733
Narcosis of permanently cationic molecules (MechoA 1.3)*	iSafeRat® daphEC50	-7.0517	-2.3188
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	iSafeRat® daphEC50	-5.0546	-2.143
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	iSafeRat® daphEC50	-5.2394	-3.3837
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® daphEC50	-6.0326	-2.2200
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® daphEC50	-5.1845	-3.3204
Redox cycling (MechoA 4.4)	iSafeRat® daphEC50	-6.7796	-5.2908
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® daphEC50	-3.8183	-2.7384
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® daphEC50	-5.723	-2.469

Table 5.3: Response domain for the different iSafeRat® algErC50 ecotoxicological local models

MechoA	Algal acute ecotoxicity models	Response domain (ErC50 in log10 (mol/L)) MIN	MAX
Non-polar narcosis (MechoA 1.1)	iSafeRat® algErC50 – Zone 1	-5.6292	-1.5631
	iSafeRat® algErC50 – Zone 2	-5.8291	-4.5799
	iSafeRat® algErC50 – Zone 3	-8	0
Polar-narcosis (MechoA 1.2)	iSafeRat® algErC50	-5.1808	-2.6792
Narcosis of permanently cationic molecules (MechoA 1.3)*	iSafeRat® algErC50	-6.8972	-2.8865
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono and poly-ester)	iSafeRat® algErC50	-6.3405	-1.4734
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® algErC50	-5.0718	-2.3838
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® algErC50	-5.7905	-3.7248
Redox cycling (MechoA 4.4)	iSafeRat® algErC50	-6.5778	-3.6141
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® algErC50	-3.5423	-3.1662
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® algErC50	-7.4350	-2.6998

** Model with Octanol-water partition coefficient as descriptor.*

c) Descriptor domain

The model is limited by its descriptor (SLWS or octanol-water partition coefficient, depending on the model) with a minimum and a maximum value which define a zone of interpolation. Beyond these extreme values, the prediction is an extrapolation. The quantified limits are given in the **Tables 6.1, 6.2 and 6.3:**

Table 6.1: Descriptor domain for the different iSafeRat® fishLC50 ecotoxicological local models

MechoA	Fish acute ecotoxicity models	Descriptor domain (SLWS or log Kow in log10 (mol/L)) MIN	MAX
Non-polar narcosis (MechoA 1.1)	iSafeRat® fishLC50 - Zone 1	-4.6292	0.8673
	iSafeRat® fishLC50 - Zone 2	Up to -5.675	Below -4.6292
	iSafeRat® fishLC50 - Zone 3	Not relevant	Below -5.675
Polar-narcosis (MechoA 1.2)	iSafeRat® fishLC50	-5.0593	0.1095
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	iSafeRat® fishLC50	-4.9909	1.6863
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	iSafeRat® fishLC50	-4.3675	0.2288
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® fishLC50	-4.2811	0.8774
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® fishLC50	-5.1256	0.4120
Redox cycling (MechoA 4.4)	iSafeRat® fishLC50	-5.0017	1.0098
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® fishLC50	-3.5528	1.9984
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® fishLC50	-5.2744	3.5692
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, simple anilines)	iSafeRat® fishLC50	-3.4683	0.2803
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, polar anilines)	iSafeRat® fishLC50	-3.5241	-0.9068

Table 6.2: Descriptor domain for the different iSafeRat® DaphEC50 ecotoxicological local models

MechoA	Daphnids acute ecotoxicity models	Descriptor domain (SLWS or log Kow in log10 (mol/L)) MIN	MAX
Non-polar narcosis (MechoA 1.1)	iSafeRat® daphEC50 – Zone 1	-4.6961	0.8673
	iSafeRat® daphEC50 – Zone 2	Up to -5.5797	Below -4.6960
	iSafeRat® daphEC50 – Zone 3	Not relevant	Below -5.5797
Polar-narcosis (MechoA 1.2)	iSafeRat® daphEC50	-5.0175	0.1095

Narcosis of permanently cationic molecules (MechoA 1.3)*	iSafeRat® daphEC50	-3.77	2.60
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	iSafeRat® daphEC50	-4.788	0.929
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	iSafeRat® daphEC50	-5.0649	-0.7654
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® daphEC50	-5.9231	0.8774
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® daphEC50	-4.7203	0.9351
Redox cycling (MechoA 4.4)	iSafeRat® daphEC50	-5.0017	1.0098
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® daphEC50	-3.5528	1.9984
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® daphEC50	-4.1631	3.7913

Table 6.3: Descriptor domain for the different iSafeRat® AlgErC50 ecotoxicological local models

MechoA	Algal acute ecotoxicological models	Descriptor domain (SLWS or log Kow in log10 (mol/L)) MIN	MAX
Non-polar narcosis (MechoA 1.1)	iSafeRat® algErC50 – Zone 1	-4.3792	0.4903
	iSafeRat® algErC50 – Zone 2	Up to -5.3650	Below -4.3792
	iSafeRat® algErC50 – Zone 3	Not relevant	Below -5.3650
Polar-narcosis (MechoA 1.2)	iSafeRat® algErC50	-3.5754	0.1095
Narcosis of permanently cationic molecules (MechoA 1.3)*	iSafeRat® algErC50	-3.92	2.60
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono and poly-ester)	iSafeRat® algErC50	-6.1467	0.929
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® algErC50	-4.9872	0.8774
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® algErC50	-4.7203	0.1386
Redox cycling (MechoA 4.4)	iSafeRat® algErC50	-4.5632	1.0098
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® algErC50	-2.3542	1.4624
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® algErC50	-4.2518	3.7913

** Model with Octanol-water partition coefficient as descriptor.*

The toxicity for a compound with SLWS outside of the descriptor range is predicted by the model but it has to be considered as an extrapolation. In that case, the prediction can still be considered as reliable if there is no contrary indication. Moreover, it recognised that there is a solubility cut-off beyond which no aquatic toxicity is observed, even for long-term exposure. This solubility cut-off value is here defined as the intersection between the solubility cut-off regression and the observed toxicity regression. The **Table 7** gives the values of the solubility cut-off:

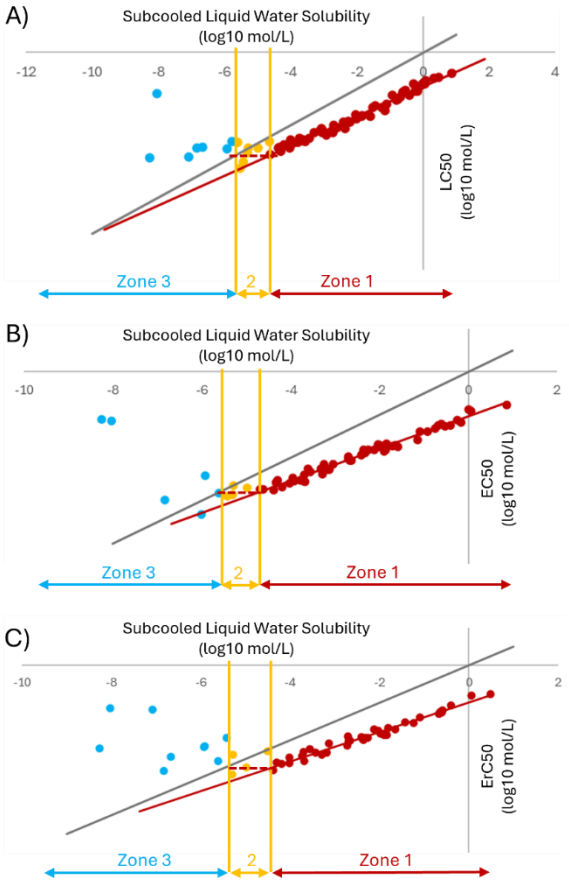
Table 7: Solubility cut-off of the iSafeRat® acute ecotoxicity models

iSafeRat® model	Solubility cut-off (log mol/L)
MechoA 1.1 – fishLC50	See figure 1 and table 6.1
MechoA 1.1 – daphEC50	See figure 1 and table 6.2
MechoA 1.1 – algErC50	See figure 1 and table 6.3
MechoA 1.2 – fishLC50	-6.491
MechoA 1.2 – daphEC50	-6.219
MechoA 1.2 – algErC50	-8.344
MechoA 1.3 – daphEC50	Not relevant
MechoA 1.3 – algErC50	Not relevant
MechoA 2.1 – fishLC50 (mono-ester)	-5.061
MechoA 2.1 – daphEC50 (mono-ester)	-5.015
MechoA 2.1 – fishLC50 (poly-ester)	-5.923
MechoA 2.1 – daphEC50 (poly-ester)	-5.260
MechoA 2.1 – algErC50 (mono and poly-ester)	-6.484
MechoA 3.1 – fishLC50	-5.607
MechoA 3.1 – daphEC50	-6.845
MechoA 3.1 – algErC50	-5.666
MechoA 3.2 – fishLC50	-5.03
MechoA 3.2 – daphEC50	-4.749
MechoA 3.2 – algErC50	-5.118
MechoA 4.4 – fishLC50	-7.557
MechoA 4.4 – daphEC50	-7.571
MechoA 4.4 – algErC50	-8.868
MechoA 5.2 – fishLC50	-3.346
MechoA 5.2 – daphEC50	-3.603
MechoA 5.2 – algErC50	-3.424
MechoA 1.2&5.2 – fishLC50	-7.297
MechoA 1.2&5.2 – daphEC50	-6.385
MechoA 1.2&5.2 – algErC50	-11.056
MechoA an4.3 & 1.2 – fishLC50 (simple anilines)	-6.042
MechoA an4.3 & 1.2 – fishLC50 (polar anilines)	-14.217

For MechoA 1.1, the experimental data for fish, daphnids and algae are composed of 3 zones within the applicability domain, as follows:

- **Zone 1** compounds fall within the standard regression far all compounds that collectively encompassed by the non-polar narcosis mechanism of toxic action (MechoA 1.1) structural domain and linear regression equation.
- **Zone 2** covers compounds where equilibrium has not been fully reached within the duration of the experiment due to their hydrophobicity and therefore no longer sit on the same regression line as zone 1 compounds but with higher effect concentrations than would be predicted for compounds in the regression used in Zone 1. Nevertheless, they still have toxicity values lower than the solubility cut-off limit and can be quantitatively estimated despite high variability between experimentally tested compound data (due to the difficulties encountered when performing studies on such hydrophobic compounds).
- **Zone 3** consists uniquely of compounds for which the experimental toxicity values were observed to be greater than their solubility limit. None of the experimental data of this zone have values which are less than the maximum solubility limit of the compound. Therefore, it is not meaningful to create a regression for this zone and all compounds falling within it can only be considered to have toxicity values greater than the maximum solubility value of the compound for these endpoints. In such cases, the solubility cut-off was determined as the point where the horizontal line in zone 2 meets the maximum solubility regression line (see table 7 and figure 1).

Figure 1: Illustration of the current three zones of applicability domain for non-polar narcosis compounds (MechoA 1.1) for acute toxicity to A) fish, B) daphnids and C) algae. The full grey line represents the water solubility cut-off, the full red line represents the model regression based on zone 1 data, the full orange lines represent the limits of the zones and the dotted horizontal red line represents the plateau used to derive an EC50 or LC50 in zone 2.



d) Structural fragment domain

The description of the structural fragments present in the molecules of the training set of each model is given in the **Tables 8.1, 8.2 and 8.3:**

Table 8.1: Structural fragments in the training set of the iSafeRat® fishLC50 acute ecotoxicity models

MechoA	Fish acute ecotoxicity models	Structural fragments
Non-polar narcosis (MechoA 1.1)	iSafeRat® fishLC50 - Zone 1	amide; lactam; ketone; acetal; 1° alcohol; 2° alcohol; 3° alcohol; benzyl alcohol; a,b-unsat. Alcohol; aromatic ether; ether; ethoxylation; mustard; chloride; fluoride ; pyridine; fused aromatic cycles; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_1; C_2; C_4; C_5; C_6; C_7; C_8; C_9; C_{10}; C_{11}; C_{12}; C_{14}; C_{15}.
	iSafeRat® fishLC50 - Zone 2	ketone; 1° alcohol; 3° alcohol; a,b-unsat. alcohol; phenyl; alkene; cycloalkyl; alkyl; C_{10}; C_{11}; C_{12}; C_{13}; C_{16}; C_{20}.
	iSafeRat® fishLC50 - Zone 3	1° alcohol; alkene; cycloalkyl; alkyl; C_{12}; C_{14}; C_{16}; C_{18}.

Polar-narcosis (MechoA 1.2)	iSafeRat® fishLC50	phenol; aromatic ether; fused aromatic cycles; phenyl; alkyl; C_6; C_7; C_8; C_{10}; C_{11}; C_{12}; C_{14}; C_{15}.
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	iSafeRat® fishLC50	triaromatic phosphate; 3° phosphate; C-aromatic ester; ester; lactone; benzyl ester; allylic ester; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; lactate; 2° alcohol; aromatic ether; ether; ethoxylation; chloride; fused aromatic cycles; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_4; C_5; C_6; C_7; C_8; C_9; C_{10}; C_{11}; C_{12}; C_{13}; C_{14}; C_{17}; C_{18}.
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	iSafeRat® fishLC50	aromatic nitro; C-aromatic ester; ester; lactone; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; ether; ethoxylation; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_4; C_6; C_7; C_8; C_{10}; C_{12}; C_{14}; C_{15}; C_{16}; C_{18}; C_{22}
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® fishLC50	ester; aldehyde; aromatic ether; epoxide; glycidyl; ether; 3° aromatic amine; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_3; C_5; C_6; C_{10}; C_{11}; C_{12}; C_{13}; C_{14}.
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® fishLC50	ester; nitrile; aldehyde; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; 2° alcohol; benzyl alcohol; aromatic ether; ethoxylation; pyridine; pseudoheteroaromatic cycle with carbonyl; pseudoheteroaromatic O; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_3; C_4; C_5; C_7; C_8; C_{11}; C_{12}; C_{14}; C_{15}; C_{33}.
Redox cycling (MechoA 4.4)	iSafeRat® fishLC50	1° alcohol; 1° thiol; alkyl; C_2; C_3; C_6; C_8; C_9.
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® fishLC50	carboxylic acid; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; lactate; 2° alcohol; aromatic ether; chloride; phenyl; alkene; alkyl; C_2; C_3; C_4; C_6; C_7; C_8; C_9; C_{10}; C_{11}.
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® fishLC50	1° amine; 2° amine; 3° amine; phenyl; cycloalkyl; alkyl; C_3; C_4; C_5; C_6; C_7; C_8; C_9; C_{12}; C_{16}; C_{20}.
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, simple anilines)*	iSafeRat® fishLC50	aromatic nitro; aromatic ketone; aromatic ether; 1° aromatic amine; 2° aromatic amine; 3° aromatic amine; phenyl; alkyl; C_6; C_7; C_8; C_9; C_{10}; C_{12}.
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, polar anilines)**	iSafeRat® fishLC50	aromatic nitro; C-aromatic ester; 1° aromatic amine; bromide; chloride; fluoride; fused aromatic cycles; phenyl; alkyl; C_6; C_9; C_{10}; C_{13}.

** Simple aniline model: primary, secondary or tertiary mono-aromatic amine, with one or several amines on the same ring, without halides or esters on this ring.*

***Polar aniline model: If there is only one aromatic amine, the ring must be substituted by at least one halide or ester. Or else, there can be several rings holding each an amine group, without the need to have halide substituents.*

Table 8.2: Structural fragments in the training set of the iSafeRat® DaphEC50 acute ecotoxicity models

MechoA	Daphnids acute ecotoxicity models	Structural fragments
Non-polar narcosis (MechoA 1.1)	iSafeRat® daphEC50 – Zone 1	amide; lactam; diaromatic ketone; aromatic ketone; ketone; acetal; 1° alcohol; 2° alcohol; 3° alcohol; benzyl alcohol; a,b-unsat. alcohol; diaromatic ether; aromatic ether; ether; ethoxylation; mustard; chloride; fluoride; pyridine; fused aromatic cycles; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_1; C_2; C_4; C_5; C_6; C_7; C_8; C_9; C_{10}; C_{11}; C_{12}; C_{13}; C_{14}; C_{15}.
	iSafeRat® daphEC50 – Zone 2	ketone; 1° alcohol; phenyl; alkene; cycloalkyl; alkyl; C_{12}; C_{13}; C_{16}.
	iSafeRat® daphEC50 – Zone 3	1° alcohol; phenyl; alkene; alkyl; C_{10}; C_{12}; C_{14}; C_{18}.
Polar-narcosis (MechoA 1.2)	iSafeRat® daphEC50	phenol; aromatic ether; phenyl; alkyl; C_6; C_7; C_8; C_9; C_{10}; C_{11}; C_{12}; C_{14}; C_{15}.
Narcosis of permanently cationic molecules (MechoA 1.3)	iSafeRat® daphEC50	a,b-unsat. carbonyl/iminyl/thionyl/nitrile; 1° alcohol; 3° aromatic amine; iminium; quaternary ammonium; phenyl; alkene; cycloalkyl; alkyl; C_5; C_{13}; C_{17}; C_{19}; C_{22}; C_{25}.
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	iSafeRat® daphEC50	3° phosphate; C-aromatic ester; ester; lactone; allylic ester; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; lactate; 1° alcohol; 2° alcohol; aromatic ether; ether; ethoxylation; chloride; fused aromatic cycles; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_4; C_5; C_6; C_7; C_8; C_9; C_{10}; C_{11}; C_{12}; C_{13}; C_{14}; C_{15}; C_{16}; C_{17}.
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	iSafeRat® daphEC50	C-aromatic ester; ester; lactone; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; ether; ethoxylation; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_8; C_{10}; C_{12}; C_{15}; C_{16}; C_{18}; C_{19}; C_{20}.
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® daphEC50	ester; aldehyde; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; aromatic ether; epoxide; glycidyl; ether; 3° aromatic amine; bromide; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_3; C_4; C_5; C_6; C_7; C_8; C_9; C_{10}; C_{11}; C_{12}; C_{13}; C_{14}; C_{15}.
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® daphEC50	ester; benzyl ester; nitrile; aldehyde; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; 1°alcohol; 2° alcohol; benzyl alcohol; aromatic ether; ether; ethoxylation; pyridine; pseudoheteroaromatic cycle with carbonyl; pseudoheteroaromatic O; non-aromatic heterocycle; phenyl; alkyne; alkene; cycloalkyl; alkyl; C_3; C_4; C_5; C_7; C_8; C_9; C_{10}; C_{11}; C_{12}; C_{16}; C_{33}.
Redox cycling (MechoA 4.4)	iSafeRat® daphEC50	1° alcohol; 1° thiol; alkyl; C_2; C_3; C_4; C_6; C_8; C_9.

Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® daphEC50	aromatic carboxylic acid; carboxylic acid; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; lactate; 2° alcohol; pyridine; alkene; alkyl; C_2; C_3; C_4; C_5; C_6; C_7; C_8; C_9; C_{10}; C_{11}.
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® daphEC50	1° amine; 2° amine; 3° amine; cycloalkyl; alkyl; C_2; C_3; C_4; C_5; C_6; C_8; C_{12}; C_{18}.

Table 8.3: Structural fragments in the training set of the iSafeRat® algErC50 acute ecotoxicity models

MechoA	Algal acute ecotoxicological models	Structural fragments
Non-polar narcosis (MechoA 1.1)	iSafeRat® algErC50 – Zone 1	amide; lactam; diaromatic ketone; aromatic ketone; ketone; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; 1° alcohol; 2° alcohol; 3° alcohol; benzyl alcohol; a,b-unsat. alcohol; aromatic ether; ether; ethoxylation; chloride; fluoride; non-aromatic heterocycle; phenyl; alkyne; alkene; cycloalkyl; alkyl;C_1; C_2; C_4 to C_{15}.
	iSafeRat® algErC50 – Zone 2	1° alcohol; phenyl; alkyl; C_{11}; C_{12}; C_{13}.
	iSafeRat® algErC50 – Zone 3	ketone; 1° alcohol; alkene; cycloalkyl; alkyl; C_{10}; C_{12}; C_{14}; C_{15}; C_{16}; C_{18}.
Polar-narcosis (MechoA 1.2)	iSafeRat® algErC50	1° aromatic amine; phenol; aromatic ether; phenyl; alkyl; C_6; C_7; C_8; C_9; C_{10}; C_{11}; C_{12}; C_{13}; C_{15}.
Narcosis of permanently cationic molecules (MechoA 1.3)	iSafeRat® algErC50	aromatic sulfonate; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; 3° aromatic amine; iminium; quaternary ammonium; phenyl; alkene; cycloalkyl; alkyl; C_4; C_{17}; C_{25}; C_{26}.
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono and poly-ester)	iSafeRat® algErC50	3° phosphate; C-aromatic ester; ester; lactone; allylic ester; ketone; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; lactate; 1° alcohol; 2° alcohol; aromatic ether; ether; ethoxylation; chloride; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_2; C_4; C_5; C_6; C_7; C_8; C_9; C_{10}; C_{11}; C_{12}; C_{13}; C_{14}; C_{15}; C_{16}; C_{17}; C_{18}; C_{19}; C_{22}.
Hard-electrophile reactivity (MechoA 3.1)	iSafeRat® algErC50	ester; aldehyde; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; aromatic ether; epoxide; glycidyl; ether; 3° aromatic amine; bromide; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_3; C_5; C_6; C_{10}; C_{11}; C_{12}; C_{13}; C_{14}; C_{15}.
Soft-electrophile reactivity (MechoA 3.2)	iSafeRat® algErC50	ester; benzyl ester; nitrile; aldehyde; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; aromatic ether; ether; ethoxylation; non-aromatic heterocycle; phenyl; alkene; cycloalkyl; alkyl; C_3; C_7; C_8; C_{10}; C_{11}; C_{12}; C_{13}; C_{15}; C_{16}.
Redox cycling (MechoA 4.4)	iSafeRat® algErC50	1° alcohol; 1° thiol; alkyl; C_2; C_4; C_8.
Acidification or alkalinization of cells (MechoA 5.2)	iSafeRat® algErC50	aromatic carboxylic acid; carboxylic acid; a,b-unsat. carbonyl/iminyl/thionyl/nitrile; pyridine; phenyl; alkene; alkyl;C_3; C_4; C_5; C_6; C_7; C_8; C_9.

Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	iSafeRat® algErC50	1° amine; 2° amine; 3° amine; non-aromatic heterocycle; phenyl; cycloalkyl; alkyl; C_3; C_4; C_5; C_6; C_7; C_8 ; C_9; C_{12} ; C_{18}.

** Most fluorinated substances are challenging to measure experimentally, both in physico-chemistry and ecotoxicology. We have decided to use only internally validated studies as K1 in our models and, exceptionally, to exclude K2 internally validated studies.*

e) Mechanistic domain

Currently, most of acute ecotoxicological models are related to one MechoA but a few are related to two MechoAs. One mechoA is represented by one or several structural alerts (Bauer et al., 2018). The iSafeRat® fishLC50, daphEC50 and algErC50 models can reliably predict the aquatic toxicity for chemicals with the MechoA:

- non-polar narcosis (MechoA 1.1)
- polar narcosis (MechoA 1.2)
- narcosis of permanently cationic molecules (MechoA 1.3, except for fish)
- mono-/poly-esters whose hydrolysis products are narcotic compounds (MechoA 2.1)
- hard-electrophile reactivity (MechoA 3.1)
- soft-electrophiles reactivity (MechoA 3.2)
- RedOx cycling compounds (MechoA 4.4)
- Acidification or alkalinization of cells (MechoA 5.2)
- Pro-reactive compounds and polar narcosis (MechoA an4.3 & 1.2, only for fish)
- Polar narcosis and acidification or alkalinization of cells (MechoA 1.2 & 5.2)

f) Metabolic domain, if relevant

Information are provided specifically to each case in QPRF.

g) Possible defined (graphical) expression of how the descriptor values of the chemicals in the training set are distributed in relation to the endpoint values predicted by the model.

The graphics of the linear regressions composing the models and the associated data are proprietary and have not been made publicly available. They have been shared with some regulatory authorities and can be shared with other regulatory authorities upon request.

5.2 Method used to assess the applicability domain

To assess if the compound under investigation falls within the applicability domain of the model, different features are evaluated:

- **1/ the mechanistic domain (section 5.1.e):**

If the compound is known to correspond to a mechanism of toxic action of an existing iSafeRat® acute ecotoxicological model, the compound is considered as falling within the mechanistic domain of the model. Otherwise, it is considered as outside the mechanistic domain and no prediction is performed.

- **2/ the descriptor domain:**

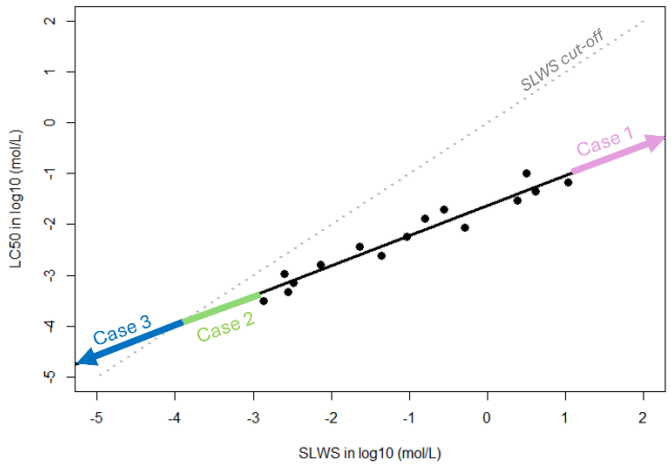
Then, the log10 SLWS (mol/L) of the compound used as model input is compared to the range of log10 SLWS (mol/L) of the compounds present in the training set of the model.

- If the descriptor value falls between the range of minimum and maximum, the test item is considered as falling within the descriptor domain of the model and the prediction is considered as an interpolation.
- If the descriptor value falls beyond the defined range of minimum and maximum values of the training set, the prediction is considered as extrapolated. However, this extrapolated status does not have the same impact on the prediction reliability, depending on the space where the descriptor value falls (**Figure 2**). Three different cases are identified and are interpreted differently:
    - **Case 1:** the SLWS value of the compound is higher than all the values of the compounds of the training set (the compound is more soluble than the ones of the training set). This is especially

the case for miscible compounds where solubility measurements have not reached a limit to dissolution within the limits of the test. For such compounds, experimental tests typically do not provide quantitative data but often use limit tests values e.g. LC50>100 mg/L. They inform us that no LC50 is reached up to the maximum concentration tested but unfortunately such data are not exploitable in linear regression. According to our model, the toxicity decreases as water solubility increases. Therefore, one can expect that even if the SLWS value of the compound is higher than the ones present in the training set, it is quite unlikely to obtain toxicity values leading to a CLP classification. Nevertheless the water solubility values may be predicted. Therefore an alternative is to use the a predicted water solubility value as descriptor to predict the ecotoxicity although this will be higher than the highest measurable descriptor value. The prediction can be considered as extrapolated but reliable with restrictions.

- **Case 2**: the descriptor value of the compound is lower than the lowest value of the descriptor and greater than the water-solubility cut-off value (except for models of MechoA 1.1). The prediction can be considered as extrapolated but reliable with restrictions depending on the extent of deviation from the lowest SLWS value of the training set.
- **Case 3:** The descriptor value falls below the water solubility cut-off value of the regression as defined in section 5.1.c. In such cases, a compound is not expected to exert intrinsic toxicity but may induce physical effects due to undissolved test compound coming into contact with the respiratory system/cell membranes of the test organism: the value is not strictly considered as an extrapolation and is considered as reliable with restrictions.

Figure 2: Graphical representation of the three different cases where a test substance falls beyond the descriptor range (example on a randomly created model).



- **3/ Response domain:**

Then, the ecotoxicological value of the compound to investigate (in log10 (mol/L)) is compared to the range of log10 ecotoxicological values (mol/L) of the compounds present in the training set of the given model. If the predicted ecotoxicological value falls within the range, the test item is considered as falling within the response domain of the model. If the predicted value falls beyond the defined range, the prediction is considered as extrapolated. Also, if the predicted value is higher than the water solubility of the test item, that means that the test item crosses the water-solubility cut-off and that the test item is not expected to exert any intrinsic toxicity.

- **4/ Structural domain:**

The molecular structure of the test compound is decomposed into structural fragments defined by KREATiS experts and compared to those present in the compounds of the training set of the model. Different cases can be identified:

- If all the fragments of the test compound are present in the compounds of the training set of the model, the compound is considered as falling within the structural applicability domain of the model.
- If one or several fragments are missing, the test item is considered as partially falling within the applicability domain. This is more or less problematic depending on the nature of the

missing fragment(s). If the missing fragments are unrelated to the MechoA of the test item, and are related to non-polar narcosis also known as baseline toxicity (e.g. alkyl, alkene, alcohol) or to another MechoA defined as less toxic in the MechoA decision tree, the prediction is still considered reliable as the missing fragments will be only interact as a contributor to hydrophobicity (which will be taken into account by the SLWS descriptor) and the prediction will still be considered as reliable.

- Alternatively, a test item may fall within the mechanistic domain but not entirely within the structural domain because too few values were found to validate the assumption and that they are not yet implemented in the models. In this case, analogues are provided to demonstrate the fit to the regression.

- **5/ Metabolic domain:**

This domain is not crucial for the iSafeRat® ecotoxicological models but if relevant, information will be provided in the related QPRF.

Synthesis of the applicability domain:

A compound falling in all domains is considered fully in the applicability domain and the prediction can be considered as reliable without restrictions.

A compound falling in some domains but not all is considered partially within the applicability domain and the prediction should be considered as an extrapolation. In case by case, this extrapolation should be considered as reliable with restrictions or unreliable, depending on the deviation(s) to the applicability domain.

A compound falling outside all domains is considered outside applicability domain and the prediction is considered as unreliable.

5.3 Software name and version for applicability domain assessment

The software iSafeRat® Desktop provides some information regarding the applicability domain. However, the full applicability domain assessment must be performed manually by the user based on the information provided in the QMRF.

5.4 Limits of applicability

The limits of applicability are defined in sections 5.1.

6 Defining goodness-of-fit and robustness (internal validation) - OECD principle 4

6.1 Availability of the training set

The training set of the model is proprietary and has not been made publicly available. The training set of the model has been shared with some regulatory authorities and can be shared with other regulatory authorities on request.

6.2 Available information for the training set

CAS RN: YES (confidential business information)
Chemical Name: YES (confidential business information)
SMILES: YES (confidential business information)
Formula: NO
INChI: NO
MOL file: NO

6.3 Data for each descriptor variable for the training set

The training set of the model is proprietary and has not been made publicly available.

6.4 Data for the dependent variable for the training set

The training set of the model is proprietary and has not been made publicly available.

6.5 Other information about the training set

The training set of the model is proprietary and has not been made publicly available.

6.6 Pre-processing of data before modelling

All the descriptor and endpoint values were converted into mol.L-1 and used as decimal logarithm for modelling purposes. The solubility values for training compounds in solid state were converted to their corresponding Subcooled Liquid Water Solubility (SLWS) values taking into account the Melting Point as an additional parameter.

All the modelling steps and goodness-of-fit and robustness calculations were performed with R software (R Core Team, 2022; version 4.2.1, 2022-06-23 ucrt). The presented metrics where obtained according to the formulae presented in the OECD document related to the validation of QSAR models (OECD, 2007). The R-script is available upon request for the competent authorities.

6.7 Statistics for goodness-of-fit

For each model, the respect of the hypothesis for a simple linear regression were investigated (residual homoscedasticity and normality).

The **Tables 9.1, 9.2 and 9.3** present the following metrics related to the goodness-of-fit of each simple linear regression: **The coefficient of multiple determination (Multiple-R^2), the adjusted coefficient of determination (Adjusted-R^2), the Root Mean Squared Error (RMSE)**.

Additionally, a significance test for linear regression was performed based on the F statistics. The H_0 assumes that there is no significant relationship between the dependent and the independent variables. If the p-value > 0.05, the hypothesis is accepted and no significant relationship occurs. If the p-value < 0.05, the hypothesis is rejected and the relationship is significant.

For the models where a significant relationship was observed (p-value < 0.05), the coefficient of determination (R^2) and the RMSE are consider as relevant indicators of a model goodness-of-fit. The closer the R^2 is to 1 and lower the RMSE value is, and the better the goodness-of-fit of the model. Low RMSE values (0.3) indicate a low deviation between the observed values of the compounds in the training set and the regression.

For a few, specific MechoA, no significant or a weak relationship between the dependent and the independent variable was observed ("p-value ≥ 0.05" or "0.05> p-value <0.01", respectively). However, the residuals (deviation between the observed and predicted values) are so low that a predicted value can be derived with confidence. That means that whatever the water solubility value, the ecotoxicological value remains similar for a defined organism and mechanism of action (i.e., hydrophobicity has little impact on toxicity). In such cases, only residuals-based metrics (e.g. RMSE) were considered as relevant to assess the validity of the model. Even if R^2 metrics do not appear relevant indicators in such cases, the values are provided for information (in italics).

If for most of models it was possible to retrieve enough data to build both training and validation sets, it was not the case for several MechoA. If less than 11 data were retrieved, all the data were used to constitute the training set of the model and enhance its goodness-ot-fit and robustness metrics. In such cases, no validation set was used, according to the Q&A document of ECHA Webinar (2024)[1]. These models having been previously shared with ECHA with their data divided between training and validation sets, metrics of the goodness of fit, robustness and predictivity of these previous versions are presented in Appendix 1, while the new versions with all data in the training set are presented in the **Tables 9.1, 9.2 and 9.3** :

[1] OECD QSAR Assessment Framework in REACH dossier evaluation: what you need to know. Events - ECHA (europa.eu). f19f111f-0896-709c-77f2-a45aa501e39f (europa.eu)

Table 9.1: Goodness-of-fit statistics for iSafeRat®fishLC50 acute ecotoxicological models

MechoA of FishLC50 models	Number of compounds (training set)	p-value	Multiple-R^2	Adjusted-R^2	RMSE
Non-polar narcosis (MechoA 1.1)	71	<2.2e-16	0.9777	0.9773	0.1754
Polar-narcosis (MechoA 1.2)	16	8.00e-10	0.9375	0.9331	0.1474
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	35	< 2.2e-16	0.9243	0.922	0.1808
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	18	8.28e-10	0.9107	0.9051	0.2050
Hard-electrophile reactivity (MechoA 3.1)	15	6.29e-06	0.8025	0.7873	0.2534
Soft-electrophile reactivity (MechoA 3.2)	*12*	*3.89e-02*	*0.3607*	*0.2967*	0.2651
Redox cycling (MechoA 4.4)	6	9.9e-04	0.9491	0.9364	0.2357
Acidification or alkalinization of cells (MechoA 5.2)	*11*	*1.76e-02*	*0.4829*	*0.4255*	0.2226
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	14	4.93e-08	0.9228	0.9164	0.3047
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, simple anilines)	12	5.98e-06	0.8818	0.87	0.2445
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, polar anilines)	14	6.73e-08	0.9188	0.912	0.1691

Table 9.2: Goodness-of-fit statistics for iSafeRat®daphEC50 acute ecotoxicological models

MechoA of DaphEC50 models	Number of compounds (training set)	p-value	Multiple-R^2	Adjusted-R^2	RMSE
Non-polar narcosis (MechoA 1.1)	61	< 2.2e-16	0.9788	0.9785	0.1464
Polar-narcosis (MechoA 1.2)	19	9.67e-11	0.9197	0.915	0.1458
Narcosis of permanently cationic molecules (MechoA 1.3)	7	9.35e-06	0.9850	0.982	0.2045
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	46	< 2.2e-16	0.9208	0.919	0.2033
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	13	5.49e-08	0.938	0.9324	0.1546
Hard-electrophile reactivity (MechoA 3.1)	30	< 2.2e-16	0.9219	0.9191	0.2599
Soft-electrophile reactivity (MechoA 3.2)	*16*	*2.32e-02*	*0.3170*	*0.2682*	0.4499

Redox cycling (MechoA 4.4)	7	6.15e-03	0.8050	0.7660	0.2462
Acidification or alkalinization of cells (MechoA 5.2)	15	1.5e-03	0.5513	0.5167	0.1791
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	16	7.01e-10	0.9387	0.9343	0.1780

Table 9.3: Goodness-of-fit statistics for iSafeRat®algErC50 acute ecotoxicological models

MechoA of AlgErC50 models	Number of compounds (training set)	p-value	Multiple-R^2	Adjusted-R^2	RMSE
Non-polar narcosis (MechoA 1.1)	45	< 2.2e-16	0.9705	0.9699	0.1705
Polar-narcosis (MechoA 1.2)	14	3.84e-08	0.926	0.9198	0.1718
Narcosis of permanently cationic molecules (MechoA 1.3)	4	3.00e-04	0.9994	0.9991	0.0393
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono and poly-ester)	49	< 2.2e-16	0.9643	0.9635	0.2009
Hard-electrophile reactivity (MechoA 3.1)	15	5.36e-09	0.9328	0.9276	0.1989
Soft-electrophile reactivity (MechoA 3.2)	*13*	*2.67e-02*	*0.3726*	*0.3156*	0.4264
Redox cycling (MechoA 4.4)	4	8.20e-03	0.9837	0.9756	0.1418
Acidification or alkalinization of cells (MechoA 5.2)	12	3.86e-01	0.0760	-0.0164	0.1194
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	15	1.65e-08	0.9201	0.914	0.3158

The data presented in the above tables were obtained using the R software. The p-value, multiple-R^2 and adjusted-R^2 metrics were obtained using the summary() function applied to the linear model created using the lm() function. The RMSE was calculated using in-house command line following the formula presented in the document OECD (2007).

6.8 Robustness - Statistics obtained by leave-one-out cross-validation

Leave-one-out (LOO) cross validation is carried out such that each compound present in the training is excluded once. A new model is built using only the remaining compounds (N-1). The ecotoxicity value of the excluded compound is then predicted using the new model. This step occurs as many time as there is data in the initial training set of the local model. Then, the **Q^2_{LOO}** (explained variance in prediction) and the **SDEP** (Standard Error in Prediction) was calculated. Ideally, the Q^2_{LOO} value approach 1 and the SDEP approach 0.

As explained in goodness-of-fit section, the Q_{2LOO} metric is not relevant for the models where no significant relationship between independent and dependent variables was observed. They were provided for information (in italic). Statistical values are presented in the **Tables 10.1, 10.2 and 10.3**:

Table 10.1: Leave-One-Out (LOO) results for the iSafeRat® fishLC50 acute ecotoxicological models

MechoA of FishLC50 models	Number of compounds (training set)	Q^2 LOO	SDEP
Non-polar narcosis (MechoA 1.1)	71	0.9764	0.1804

Polar-narcosis (MechoA 1.2)	16	0.9020	0.1846
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	35	0.9136	0.1931
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	18	0.8896	0.2279
Hard-electrophile reactivity (MechoA 3.1)	15	0.7388	0.2914
Soft-electrophile reactivity (MechoA 3.2)	*12*	*0.0033*	*0.3310*
Redox cycling (MechoA 4.4)	6	0.8675	0.3803
Acidification or alkalinization of cells (MechoA 5.2)	*11*	*0.2046*	*0.2761*
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	14	0.8903	0.3637
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, simple anilines)	12	0.836	0.2875
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, polar anilines)	14	0.8993	0.1883

Table 10.2: Leave-One-Out (LOO) results for the iSafeRat® daphEC50 acute ecotoxicological models

MechoA of DaphEC50 models	Number of compounds (training set)	Q^2 LOO	SDEP
Non-polar narcosis (MechoA 1.1)	61	0.9772	0.1518
Polar-narcosis (MechoA 1.2)	19	0.9021	0.1609
Narcosis of permanently cationic molecules (MechoA 1.3)	7	0.9662	0.3075
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	46	0.9128	0.2132
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	13	0.9042	0.1923
Hard-electrophile reactivity (MechoA 3.1)	30	0.9093	0.2801
Soft-electrophile reactivity (MechoA 3.2)	*16*	*0.0984*	*0.5168*
Redox cycling (MechoA 4.4)	7	0.6643	0.3231
Acidification or alkalinization of cells (MechoA 5.2)	15	0.3895	0.2088
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	16	0.9035	0.2234

Table 10.3: Leave-One-Out (LOO) results for the iSafeRat® algErC50 acute ecotoxicological models

MechoA of AlgErC50 models	Number of compounds (training set)	Q^2 LOO	SDEP
Non-polar narcosis (MechoA 1.1)	45	0.9677	0.1785
Polar-narcosis (MechoA 1.2)	14	0.8935	0.2060
Narcosis of permanently cationic molecules (MechoA 1.3)	4	-	-
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono and poly-ester)	49	0.9620	0.2073
Hard-electrophile reactivity (MechoA 3.1)	15	0.9101	0.2300
Soft-electrophile reactivity (MechoA 3.2)	*13*	*0.0631*	*0.5211*
Redox cycling (MechoA 4.4)	4	-	-
Acidification or alkalinization of cells (MechoA 5.2)	*12*	*-0.2947*	*0.1413*
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	15	0.8593	0.4192

The data presented in the above tables were obtained using the R software. The metrics Q^2LOO were obtained from using the LOO() function of the DEMOVA package (version 1.0, Prana (2016)) applied to the training set of the local model. The SDEP was calculated using in-house command line based on the LOO() function result, following the formula presented in the document OECD (2007).

6.9 Robustness - Statistics obtained by leave-many-out cross-validation

Leave-Many-Out (LMO) cross validation is similar to Leave-one-out concept, however multiple training compounds are excluded from the training set and the remaining training set is used to predict the excluded compounds. The number of block of substance to exclude (iterations) was defined in order that the training set of the model created at each iteration contained ca. 75% of the original training set. Ideally, the **Q^2_{LMO}** and **SDEP** values approach 1 and 0, respectively.

As explained in goodness-of-fit section, the **Q^2_{LMO}** metric is not relevant for the models where no significant relationship between independent and dependent variables was observed. They were provided for information (data in italic). Statistical values are presented in the **Tables 11.1, 11.2 and 11.3**:

Table 11.1: Leave-Many-Out (LMO) results for the iSafeRat® fishLC50 acute ecotoxicological models

MechoA of FishLC50 models	Number of compounds (training set)	Q^2 LMO	SDEP
Non-polar narcosis (MechoA 1.1)	71	0.9772	0.1772
Polar-narcosis (MechoA 1.2)	16	0.8817	0.2029
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	35	0.9146	0.1920
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	18	0.8753	0.2423
Hard-electrophile reactivity (MechoA 3.1)	15	0.7285	0.2971

Soft-electrophile reactivity (MechoA 3.2)	*12*	*-0.2716*	*0.3738*
Redox cycling (MechoA 4.4)	6	0.8712	0.3751
Acidification or alkalinization of cells (MechoA 5.2)	*11*	*-0.7300*	*0.3063*
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	14	0.8685	0.3979
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, simple anilines)	12	0.8187	0.3027
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, polar anilines)	14	0.9036	0.1843

Table 11.2: Leave-Many-Out (LMO) results for the iSafeRat® daphEC50 acute ecotoxicological models

MechoA of DaphEC50 models	Number of compounds (training set)	Q^2 LMO	SDEP
Non-polar narcosis (MechoA 1.1)	61	0.9779	0.1496
Polar-narcosis (MechoA 1.2)	19	0.9115	0.1531
Narcosis of permanently cationic molecules (MechoA 1.3)	7	-	-
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	46	0.9100	0.2167
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	13	0.8803	0.2149
Hard-electrophile reactivity (MechoA 3.1)	30	0.9104	0.2783
Soft-electrophile reactivity (MechoA 3.2)	*16*	*-0.3919*	*0.6423*
Redox cycling (MechoA 4.4)	7	0.6653	0.3226
Acidification or alkalinization of cells (MechoA 5.2)	15	0.3691	0.2123
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	16	0.8961	0.2318

Table 9.3: Leave-Many-Out (LMO) results for the iSafeRat® algErC50 acute ecotoxicological models

MechoA of AlgErC50 models	Number of compounds (training set)	Q^2 LMO	SDEP
Non-polar narcosis (MechoA 1.1)	45	0.9651	0.1855
Polar-narcosis (MechoA 1.2)	14	0.8738	0.2243
Narcosis of permanently cationic molecules (MechoA 1.3)	4	-	-
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono and poly-ester)	49	0.9616	0.2083

Hard-electrophile reactivity (MechoA 3.1)	15	0.9054	0.2359
Soft-electrophile reactivity (MechoA 3.2)	*13*	*0.0121*	*0.5352*
Redox cycling (MechoA 4.4)	4	-	-
Acidification or alkalinization of cells (MechoA 5.2)	*12*	*-0.3333*	*0.1434*
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	15	0.8351	0.4538

The data presented in the above tables were obtained using the R software. The metrics Q^2LMO were obtained from using the LMO() function of the DEMOVA package (version 1.0, Prana (2016)) applied to the training set of the local model. The SDEP was calculated using in-house command line.

6.10 Robustness - Statistics obtained by Y-scrambling

This validation approach consists in identifying if the correlation was based on chance. The ecotoxicological data are randomly permuted and attributed to another compound of the training set. A new regression is built and its respective multiple-R^2 value is calculated. This operation is performed a certain number of times. Finally, the mean of all the R^2 obtained among the different iterations and its standard deviation (**SD**) is computed. It appears not relevant to use a number of iterations higher than the maximal number of possible permutations in the training set (factorial number) as in this case, a same regression will occur several times. Therefore, the number of iterations was set at 300 (factorial number for a training set of 6 substances in the training set). Statistical values are presented in the **Tables 12.1, 12.2 and 12.3**:

Table 12.1: Y-scrambling results for the iSafeRat® fishLC50 acute ecotoxicological models

MechoA of FishLC50 models	Number of compounds (training set)	Y-scrambling multiple-R^2	Y-scrambling SD
Non-polar narcosis (MechoA 1.1)	71	0.0140	0.0208
Polar-narcosis (MechoA 1.2)	16	0.0749	0.0934
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	35	0.0324	0.0422
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	18	0.0666	0.0778
Hard-electrophile reactivity (MechoA 3.1)	15	0.0785	0.0995
Soft-electrophile reactivity (MechoA 3.2)	*12*	*0.0828*	*0.1048*
Redox cycling (MechoA 4.4)	6	0.1989	0.2105
Acidification or alkalinization of cells (MechoA 5.2)	*11*	*0.0979*	*0.1143*
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	14	0.0772	0.0993
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, simple anilines)	12	0.097	0.1144

metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, polar anilines)	14	0.0888	0.1155

Table 12.2: Y-scrambling results for the iSafeRat® daphEC50 acute ecotoxicological models

MechoA of DaphEC50 models	Number of compounds (training set)	Y-scrambling multiple-R^2	Y-scrambling SD
Non-polar narcosis (MechoA 1.1)	61	0.0150	0.0204
Polar-narcosis (MechoA 1.2)	19	0.0630	0.0769
Narcosis of permanently cationic molecules (MechoA 1.3)	7	0.1633	0.1936
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	46	0.0203	0.0295
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	13	0.0856	0.1019
Hard-electrophile reactivity (MechoA 3.1)	30	0.0383	0.0590
Soft-electrophile reactivity (MechoA 3.2)	*16*	*0.0757*	*0.1074*
Redox cycling (MechoA 4.4)	7	0.1457	0.1730
Acidification or alkalinization of cells (MechoA 5.2)	15	0.0639	0.0815
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	16	0.0721	0.1279

Table 12.3: Y-scrambling results for the iSafeRat® algErC50 acute ecotoxicological models

MechoA of AlgErC50 models	Number of compounds (training set)	Y-scrambling multiple-R^2	Y-scrambling SD
Non-polar narcosis (MechoA 1.1)	45	0.0217	0.0288
Polar-narcosis (MechoA 1.2)	14	0.0808	0.1010
Narcosis of permanently cationic molecules (MechoA 1.3)	4	-	-
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono and poly-ester)	49	0.0201	0.0275
Hard-electrophile reactivity (MechoA 3.1)	15	0.0670	0.0948
Soft-electrophile reactivity (MechoA 3.2)	*13*	*0.0875*	*0.1146*
Redox cycling (MechoA 4.4)	4	-	-
Acidification or alkalinization of cells (MechoA 5.2)	*12*	*0.0889*	*0.1153*
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	15	0.0734	0.1114

The data presented in the above tables were obtained using the R software. The metrics Y-scrambling multiple-R^2 and Y-scrambling SD were obtained using the scramb() function of the DEMOVA package (version 1.0, Prana (2016)) applied to the training set of the local model.

6.11 Robustness - Statistics obtained by bootstrap

Bootstrapping is another form of internal validation in which a selected subset of the original training set forms the model. Each of the samples of this subset are reiterated in the model such that the newly created model is of the same size as the original training set and the excluded training samples form the validation set and predictions for the validation set are recorded each time. To obtain a Bootstrap-Q^2 value, the training set was randomly sampled at each iteration with a size representing 75% of the training set. Predictions were performed on the 25% excluded from the training set. Based on the derived predictions, a Q^2 was calculated at each iteration. The results provided in the **Tables 13.1, 13.2 and 13.3** were derived from a bootstrap with 1000 iterations.

Table 13.1: Bootstrap results for the iSafeRat® fishLC50 acute ecotoxicological models

MechoA of FishLC50 models	Number of compounds (training set)	Bootstrap-Q^2	Bootstrap-SDEP
Non-polar narcosis (MechoA 1.1)	71	0.9754	0.1839
Polar-narcosis (MechoA 1.2)	16	0.8812	0.1970
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	35	0.8900	0.2030
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	18	0.8700	0.2441
Hard-electrophile reactivity (MechoA 3.1)	15	0.6284	0.3205
Soft-electrophile reactivity (MechoA 3.2)	*12*	*-0.2532*	*0.3482*
Redox cycling (MechoA 4.4)	6	-	-
Acidification or alkalinization of cells (MechoA 5.2)	*11*	*-0.3902*	*0.3118*
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	14	0.8120	0.4348
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, simple anilines)	12	0.7359	0.3132
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, polar anilines)	14	0.8247	0.2195

Table 13.2: Bootstrap results for the iSafeRat® daphEC50 acute ecotoxicological models

MechoA of DaphEC50 models	Number of compounds (training set)	Bootstrap-Q^2	Bootstrap-SDEP
Non-polar narcosis (MechoA 1.1)	61	0.9757	0.1562
Polar-narcosis (MechoA 1.2)	19	0.8666	0.1731
Narcosis of permanently cationic molecules	7	-	-

(MechoA 1.3)			
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	46	0.9048	0.2192
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	13	0.8712	0.2151
Hard-electrophile reactivity (MechoA 3.1)	30	0.8982	02903
Soft-electrophile reactivity (MechoA 3.2)	*16*	*-0.3917*	*0.5699*
Redox cycling (MechoA 4.4)	7	0.1504	0.4655
Acidification or alkalinization of cells (MechoA 5.2)	15	-0.3428	0.2261
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	16	0.7673	0.2463

Table 13.3: Bootstrap results for the iSafeRat® algErC50 acute ecotoxicological models

MechoA of AlgErC50 models	Number of compounds (training set)	Bootstrap-Q^2	Bootstrap-SDEP
Non-polar narcosis (MechoA 1.1)	45	0.9634	0.1849
Polar-narcosis (MechoA 1.2)	14	0.8252	0.2319
Narcosis of permanently cationic molecules (MechoA 1.3)	4	-	-
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono and poly-ester)	49	0.9562	0.2126
Hard-electrophile reactivity (MechoA 3.1)	15	0.8529	0.2616
Soft-electrophile reactivity (MechoA 3.2)	13	*-0.8033*	*0.5995*
Redox cycling (MechoA 4.4)	4	-	-
Acidification or alkalinization of cells (MechoA 5.2)	12	*-0.5340*	*0.1572*
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	15	0.6259	0.4313

The data presented in the above tables were obtained using the R software. The bootstrap Q^2 and SDEP were obtained using an in-house script following the guideline provided in the OECD (2007) document.

6.12 Robustness - Statistics obtained by other methods

No additional statistical methods were implemented.

7 Defining predictivity (External validation) - OECD Principle 4

7.1 Availability of the external validation set

The validation set of the model is proprietary and has not been made publicly available. The validation set of the model has been shared with some regulatory authorities and can be shared with other regulatory authorities on request.

7.2 Available information for the external validation set

CAS RN: YES (confidential business information)
Chemical Name: YES (confidential business information)
SMILES: YES (confidential business information)
Formula: NO
INChI: NO
MOL file: NO

7.3 Data for each descriptor variable for the external validation set

The validation set of the model is proprietary and has not been made publicly available.

7.4 Data for the dependent variable for the external validation set

The validation set of the model is proprietary and has not been made publicly available.

7.5 Other information about the external validation set

The validation set of the model is proprietary and has not been made publicly available.

7.6 Experimental design of test set

The test set of compounds were retrieved from one of the following publicly available data resources:

a) EPISuite package (available experimental results)
b) ECHA dissemination database
c) Data from KREATiS inventory

Test set selection had been a crucial exercise to demonstrate the validity of the presented model. It was made sure that the measured values were derived under appropriate experimental conditions and using suitable methods. To qualify as a test set compound, the following verification checks were made:

a) should not be a part of the training set for the model
b) should be within the applicability domain of the model
c) must have experimentally derived values available

7.7 Predictivity - Statistics obtained by external validation

The external validation was performed on a set of compounds which are not present in the training set of the models. The ecotoxicity of each compound present in the external validation set of the models was predicted. The ecotoxicological predicted values were then used to calculate the **externally explained variance (Q^2 ext)** and **Standard Deviation Error of Prediction (SDEP)**. In theory, Q^2 values closer to 1 and SDEP closer to 0 indicates that the model is associated with a reliable predictivity. **Tables 14.1, 14.2** and **14.3** provide the results derived applying the iSafeRat® models on the external validation set for each model.

For some models, it was complicated to find enough reliable data to comfortably constitute both a training and a validation set. In such cases, it was decided that the model would be more accurate if all the data were used in the training set to obtain the most robust model. Therefore, it was not possible to compute an external Q^2 for now (although efforts are continuously made to increase the number of quality data in our models). Validation set is especially of importance to detect model overfitting and for models based on advanced statistical approaches and/or based on many descriptors. KREATiS QSARs are based on mechanistic understanding and

one descriptor. Therefore, these models are less susceptible to overfitting and recent information clarified that the EU Chemicals Agency, ECHA accepts exclusion of the external validation set for such cases[2].

Table 14.1: External validation results for the iSafeRat® fishLC50 acute ecotoxicological models

MechoA of FishLC50 models	Number of compounds (training set)	Number of compounds (validation set)	External Q^2	SDEP
Non-polar narcosis (MechoA 1.1)	71	22	0.9303	0.2395
Polar-narcosis (MechoA 1.2)	16	5	0.7073	0.1592
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	35	10	0.8552	0.2210
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	18	5	0.8277	0.2718
Hard-electrophile reactivity (MechoA 3.1)	15	8	0.7166	0.2186
Soft-electrophile reactivity (MechoA 3.2)	12	6	*0.5554*	0.2223
Redox cycling (MechoA 4.4)	6	-	-	-
Acidification or alkalinization of cells (MechoA 5.2)	11	5	*0.3028*	0.2650
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	14	7	0.9310	0.3078
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, simple anilines)	12	5	0.7676	0.1956
metabolism into nitroso, hydroxylamine or nitrenium, generating DNA adducts, oxidative stress and carcinogenicity for animals & polar narcosis for all species (MechoA an4.3 & 1.2, polar anilines)	14	-	-	-

Table 14.2: External validation results for the iSafeRat® daphEC50 acute ecotoxicological models

MechoA of DaphEC50 models	Number of compounds (training set)	Number of compounds (validation set)	External Q^2	SDEP
Non-polar narcosis (MechoA 1.1)	61	23	0.8417	0.3453
Polar-narcosis (MechoA 1.2)	19	5	0.7158	0.2204
Narcosis of permanently cationic molecules (MechoA 1.3)	7	-	-	-
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono-ester)	46	12	0.8907	0.2276
Hydrolysis to non-polar narcotic products (MechoA 2.1, poly-ester)	13	-	-	-
Hard-electrophile reactivity (MechoA 3.1)	30	9	0.7848	0.2948

[2] f19f111f-0896-709c-77f2-a45aa501e39f (europa.eu)

Soft-electrophile reactivity (MechoA 3.2)	16	8	*0.3476*	0.3965
Redox cycling (MechoA 4.4)	7	-	-	-
Acidification or alkalinization of cells (MechoA 5.2)	15	-	-	-
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	16	-	-	-

Table 14.3: External validation results for the iSafeRat® algErC50 acute ecotoxicological models

MechoA of AlgErC50 models	Number of compounds (training set)	Number of compounds (validation set)	External Q^2	SDEP
Non-polar narcosis (MechoA 1.1)	45	24	0.8916	0.2969
Polar-narcosis (MechoA 1.2)	14	-	-	-
Narcosis of permanently cationic molecules (MechoA 1.3)	4	-	-	-
Hydrolysis to non-polar narcotic products (MechoA 2.1, mono and poly-ester)	49	19	0.8974	0.2976
Hard-electrophile reactivity (MechoA 3.1)	15	7	0.8980	0.2261
Soft-electrophile reactivity (MechoA 3.2)	13	5	*0.4976*	0.3915
Redox cycling (MechoA 4.4)	4	-	-	-
Acidification or alkalinization of cells (MechoA 5.2)	12	-	-	-
Polar narcosis & alkalinization of cells, corrosion if directly applied, for all species (MechoA 1.2 & 5.2)	15	6	0.8419	0.2882

The data presented in the above tables were obtained using the R software. The metrics external Q^2 and SDEP were obtained using an in-house script following the guideline provided in the OECD (2007) document.

7.8 Predictivity - Assessment of the external validation set

All attempts were made to ensure that the validation set covered the chemical domain of the training data sufficiently to avoid any possible bias in the resulting statistical validation.

7.9 Comments on the external validation of the model

The validation set may be extended from time to time and the revised validation results can be presented as the updated version of this QMRF.

8 Providing a mechanistic interpretation - OECD Principle 5

8.1 Mechanistic basis of the model

The thermodynamic relationship between surrogates for chemical activity, such as SLWS, log K_{OW} and narcosis has been widely reported in the literature (Mackay *et al.*, 2009) but only recently was it postulated that this could be used to develop a method which can be applied to certain mechanisms of action to reliably predict endpoint values for standard regulatory guideline studies (ECETOC, 2013; Thomas *et al.,* 2015, 2019).

8.2 A priori or a posteriori mechanistic interpretation

A priori

8.3 Other information about the mechanistic interpretation

To allow a better understanding of the methodology, no inexplicable molecular descriptors or modelling algorithms were included. As hypothesised by Mackay *et al.* (2009) and validated by Thomas *et al.* (2015), aquatic toxicity of a non-polar narcotic (MechoA 1.1) compound is directly correlated to its chemical activity and thereby its SLWS. For other MechoAs, specific relationships with SLWS have also been determined by KREATiS although the slopes and intercepts vary from one MechoA to another. The toxicity of MechoA 1.1 compounds can be considered as a membrane saturation phenomenon. Reasons for excess toxicity for other mechanisms of action have been elucidated in Bauer et al. (2018) and more recently is Firman et al (2022) and when they are known they are indicated in the HA-QSAR study report accompanying the QMRF.

9 Miscellaneous information

9.1 Comments

This QMRF can be used as a reference document for QPRF providing:

- iSafeRat® fishLC50 v2.1 predictions for ACUTE TOXICITY TO FISH (96-HOURS LC50);
- iSafeRat® daphEC50 v2.1 predictions for ACUTE TOXICITY TO DAPHNIDS (48-HOUR EC50);
- iSafeRat® algErC50 v2.2 predictions for TOXICITY TO ALGAE (72-HOUR ErC50).

Finally, it is important to explain that KREATiS designates a QSAR as a High Accuracy QSAR or HA-QSAR when the model meets the following internally decided obligations:

1) the model is fully compliant with the 5 OECD Principles of QSAR development;
2) that the input data have been carefully checked internally to verify fitness for purpose and compliance with the OECD guidelines. This goes beyond meeting the Klimisch score applied in an ECHA dossier, for example. All data are revalidated internally according to KREATiS criteria often leading to a more restrictive classification than the ECHA Klimisch score;
3) All the HA-QSAR models are associated with statistics highlighting high model robustness like for example high R^2 and/or low RMSE. Such results are observed in the cases where the toxicity is highly dependent upon the descriptor (e.g., MechoA 1.1 and 2.1) and those models can be considered "HA-QSAR" models without reservation;
4) Conversely, for certain MechoA subclasses, there is a less clear relationship, or even no apparent link between the hydrophobicity descriptor (i.e. water solubility) and the predicted variable, resulting in a horizontal (or near horizontal) regression (cf. p-values in goodness-of-fit tables). However, the fact that the data are close to the regression line and are not dispersed in a heterogeneous manner indicate that whatever the water solubility, the toxicity remains in a specific range. This may be due to a direct reaction between the compound and the biological matrix causing significant damage to the organism in question regardless of the water solubility of the compound until reaching the water solubility-toxicity cut-off line. This can result in low R^2 because of the poor relationship between solubility and toxicity. Nevertheless, the RMSE may still be low (potentially even lower than in HA-QSARs) as the toxicity is still dependent on the primary variable which is the MechoA. In these cases the models may still be considered as HA-QSARs as long as the RMSE is low enough (i.e. <0.3).

However, if the R^2 and RMSE limits are not met and the model does not have necessarily enough data to compute all the robustness statistics, the models do not receive the KREATiS "HA-QSAR" designation.

9.2 Bibliography

Bauer, F.J., Thomas, P.C., Fouchard, S.Y., and Neunlist, S.J.M. (2018). High-accuracy prediction of mechanisms of action using structural alerts. Computational Toxicology 7, 36–45.

ECETOC (2013). Activity-Based Relationships for Aquatic Ecotoxicology Data: Use of the Activity Approach to Strengthen MoA Predictions. (Brussels: Elsevier), pp. 547–548.

European Chemicals Agency (2016). Guidance on information requirements and chemical safety assessment Chapter R.7a: Endpoint specific guidance. ECHA, Helsinki.

Firman, J., Ebbrell, D., Bauer, F., Sapounidou, M., Hodges, G., Campos, B., Roberts, J., Gutsell, S., Thomas, P., Bonnell, M., Cronin, M. (2022). Construction of an In Silico Structural Profiling Tool Facilitating Mechanistically Grounded Classification of Aquatic Toxicants Environmental Science & Technology 2022 56 (24), 17805-17814

Letinski, D.J., Connelly, M.J., Peterson, D.R., Parkerton, T.F. (2002). Slow-stir water solubility measurements of selected alcohols and diesters. Chemosphere 48, 257–265.

Mackay, D., Arnot, J.A., Petkova, E.P., Wallace, K.B., Call, D.J., Brooke, L.T., and Veith, G.D. (2009). The physicochemical basis of QSARs for baseline toxicity. SAR and QSAR in Environmental Research 20, 393–414.

OECD (1995a). Test No. 105: Water Solubility, OECD Guidelines for the Testing of Chemicals, Section 1. OECD. https://doi.org/10.1787/9789264069589-en

OECD (1995b). Test No. 107: Partition Coefficient (n-octanol/water): Shake Flask Method. Organisation for Economic Co-operation and Development, Paris.

OECD (2004). Test No. 202: Daphnia sp. Acute Immobilisation Test, OECD Guidelines for the Testing of Chemicals, Section 2, Éditions OCDE, Paris, https://doi.org/10.1787/9789264069947-en.

OECD (2007). GUIDANCE DOCUMENT ON THE VALIDATION OF (QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIP [(Q)SAR] MODELS. OECD Environment Health and Safety Publications Series on Testing and Assessment No. 69.

OECD (2011). Test No. 201: Freshwater Alga and Cyanobacteria, Growth Inhibition Test, OECD Guidelines for the Testing of Chemicals, Section 2, Éditions OCDE, Paris, https://doi.org/10.1787/9789264069923-en.

OECD (2013), Test No. 236: Fish Embryo Acute Toxicity (FET) Test, OECD Guidelines for the Testing of Chemicals, Section 2, Éditions OCDE, Paris, https://doi.org/10.1787/9789264203709-en.

OECD (2019). Test No. 203: Fish, Acute Toxicity Test, OECD Guidelines for the Testing of Chemicals, Section 2, Éditions OCDE, Paris, https://doi.org/10.1787/9789264069961-en.

OECD (2022a). Test No. 123: Partition Coefficient (1-Octanol/Water): Slow-Stirring Method. Organisation for Economic Co-operation and Development, Paris.

OECD (2022b). Test No. 117: Partition Coefficient (n-octanol/water), HPLC Method, OECD Guidelines for the Testing of Chemicals, Section 1. OECD. https://doi.org/10.1787/9789264069824-en.

Prana, V. (2016). DEMOVA: DEvelopment (of Multi-Linear QSPR/QSAR) MOdels VAlidated using Test Set_. R package version 1.0, <https://CRAN.R-project.org/package=DEMOVA>.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Thomas, P., Dawick, J., Lampi, M., Lemaire, P., Presow, S., van Egmond, R., Arnot, J.A., Mackay, D., Mayer, P., and Galay Burgos, M. (2015). Application of the Activity Framework for Assessing Aquatic Ecotoxicology Data for Organic Chemicals. Environmental Science & Technology 49, 12289–12296.

Thomas, P.C., Bicherel, P., and Bauer, F.J. (2019). How in silico and QSAR approaches can increase confidence in environmental hazard and risk assessment: How QSAR Models Can Improve Environmental Risk Assessment. Integrated Environmental Assessment and Management 15, 40-50.

Yalkowsky, S.H. (1979). Estimation of Entropies of Fusion of Organic Compounds. Industrial & Engineering chemistry Fundamentals 18, 108–111. https://doi.org/10.1021/i160070a003

Supporting information
Training set(s):
Proprietary. All the queries must be directly addressed to KREATiS SAS.

Test set(s):
Proprietary. All the queries must be directly addressed to KREATiS SAS.

Supporting information:
None

10 Summary (KREATiS QMRF Database)

10.1 QMRF number (For KREATiS internal records only)

KTS/QMRF/ACU/05

10.2 Publication date

16 June 2025

10.3 Keywords

iSafeRat®; HA-QSAR; QSAR; Ecotoxicity; acute; fish; daphnid; algae; mortality; immobility; growth

10.4 Comments

None

Appendix 1: Applicability domains and statistical metrics of models before the training set and the validation set were merged.

These tables are related to the models which have a limited number of reliable data and for which the entire set of data was, in the previous version (version 1.11 of iSafeRat® fishLC50, daphEC50 and algErC50), splatted into a training and a validation set. In the updated model version (v2.0), the training set and the validation set were merged in a single training set to obtain more robust models. More information about the previous version of the models is available in QMRF v2.3.

Appendix-Table 1: Descriptor, response and mechanistic domains for iSafeRat® Ecotoxicity models

Models	Mechanistic domain	Reliable Descriptor range for subcooled liquids *(log10 SLWS mol/L)*	Reliable Prediction range *(log10 EC10 mol/L)*
Short-term toxicity to daphnia (poly-esters)	MechoA 2.1	-5.0649 to -0.7654	-5.2394 to -3.3837
Short-term toxicity to algae	MechoA 1.2	-3.5754 to 0.1095	-5.1808 to -2.6792
Short-term toxicity to daphnia	MechoA 1.2&5.2	-4.1631 to 3.7913	-5.723 to -2.469
Short-term toxicity to daphnia	MechoA 5.2	-3.5528 to 1.9984	-2.7384 to -3.708
Short-term toxicity to algae		-2.2619 to 1.4624	-3.5182 to -3.1662
Acute toxicity to fish (polar anilines)	MechoA an1.2&4.3	-3.5241 to -0.9068	-5.6443 to -3.5089

Appendix-Table 2: Goodness-of-fit statistical values for iSafeRat® Ecotoxicity models

iSafeRat® models	n	R^2	RMSE
Short-term toxicity to daphnia (poly-esters)	8	0.9431	0.1701
Short-term toxicity to algae (alkyl/alkoxyphenols)	8	0.9605	0.1673
Short-term toxicity to daphnia (amines)	11	0.9602	0.1631
Short-term toxicity to daphnia (carboxylic acids)	10	0.6924	0.1768
Short-term toxicity to algae (carboxylic acids)	7	0.0024	0.1133
Acute toxicity to fish (polar anilines)	10	0.9113	0.1864

Appendix-Table 3: Leave-One-Out cross validation for iSafeRat® Ecotoxicity models

iSafeRat® models	n	Q^2_{LOO}
Short-term toxicity to daphnia (poly-esters)	8	0.8934
Short-term toxicity to algae (alkyl/alkoxyphenols)	8	0.9116
Short-term toxicity to daphnia (amines)	11	0.9490
Short-term toxicity to daphnia (carboxylic acids)	10	0.5698
Short-term toxicity to algae (carboxylic acids)	7	-1.1433
Acute toxicity to fish (polar anilines)	10	0.7975

Appendix-Table 4: Predictivity statistics for iSafeRat® Ecotoxicity models

iSafeRat® models	n	n_{Test}	Q^{2*}	RMSEP*
Short-term toxicity to daphnia (poly-esters)	8	5	0.9719	0.1791
Short-term to algae (alkyl/alkoxyphenols)	8	5	0.8642	0.2155
Short-term toxicity to daphnia (amines)	11	5	0.7748	0.2117
Short-term toxicity to daphnia (carboxylic acids)	10	5	0.0431	0.2046
Short-term toxicity to algae (carboxylic acids)	7	5	0.2507	0.1347
Acute toxicity to fish (polar anilines)	10	4	0.9984	0.1387