Joint Research Programme
BTO 2023.015 | February 2023

Evaluation of QSAR tools in combination with bioassays for transformation products and emerging substances

Joint Research Programme

KWR

Bridging Science to Practice

Report

Evaluation of QSAR tools in combination with bioassays for transformation products and emerging substances

BTO 2023.015 | February 2023

This research is part of the Joint Research Programme of KWR, the water utilities and Vewin.

Project number
402045/152

Project manager
Dr. Patrick Bauerlein

Client
BTO - Thematical research - Chemical safety

Author(s)
Astrid Reus, MSc., Dr. Renske Hoondert, Dr. Miina Yanagihara

Quality Assurance
Dr. Milou Dingemans

Sent to
This report is distributed to BTO-participants.

A year after publication it is public.

Keywords

In silico toxicology, QSAR, read-across, bioassay, mutagenicity, Ames fluctuation test

Year of publishing
2023

More information
Astrid Reus
T +3130 606 9703
E Astrid.Reus@kwrwater.nl

PO Box 1072
3430 BB Nieuwegein
The Netherlands

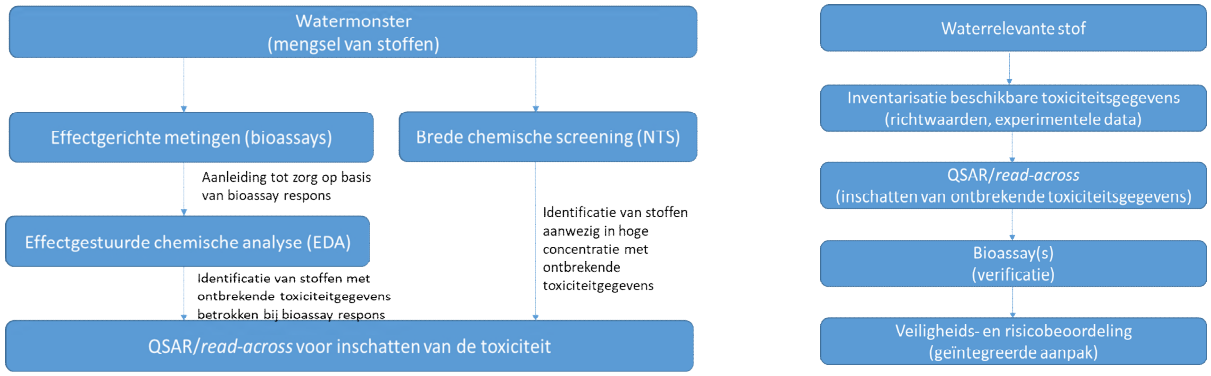
T +31 (0)30 60 69 511
E info@kwrwater.nl
I www.kwrwater.nl

Managementsamenvatting

Evaluatie van QSAR tools in combinatie met bioassays voor transformatieproducten en opkomende stoffen

Auteurs Astrid Reus, MSc., dr. Renske Hoondert, dr. Miina Yanagihara

De drinkwatersector kent vele uitdagingen voor wat betreft de waterkwaliteit van hun bronnen als gevolg van chemische bedreigingen. Door een gebrek aan toxicologische informatie voor veel opkomende stoffen is het niet altijd mogelijk om mogelijke risico's voor de volksgezondheid en het milieu te beoordelen. *In silico* (computer)tools voor het voorspellen van de toxiciteit van stoffen bieden hiervoor een oplossing. Op basis van de chemische structuur kunnen de mogelijke schadelijke effecten van een stof worden beoordeeld met behulp van kwantitatieve structuur-activiteitsrelaties (*quantitative structure-activity relationships,* QSAR) of read-across naar nauw verwante stoffen. Uit ons onderzoek blijkt dat methoden met bacteriën of cellen (bioassays) kunnen worden gebruikt om de met *in silico*-tools verkregen resultaten te verifiëren. Deze informatie biedt perspectieven voor de preventie van risico's voor de menselijke gezondheid en het milieu.



Geïntegreerde teststrategieën voor de beoordeling van watermonsters (links) en afzonderlijke waterrelevante stoffen (rechts)

Belang: beoordeling van potentiële risico's van stoffen in water

In drinkwaterbronnen kunnen lage concentraties van veel verschillende stoffen met uiteenlopende fysisch-chemische eigenschappen worden aangetroffen. Voor nieuwe, opkomende stoffen is er vaak weinig informatie beschikbaar over de toxische eigenschappen en de mogelijk schadelijke effecten op de volksgezondheid en het milieu. In een ongewone situatie of een noodgeval, wanneer water plotseling een hoge concentratie van een chemische stof bevat, is het van essentieel belang om snel de potentiële risico's voor de drinkwaterproductie, de volksgezondheid en het milieu te beoordelen.
Een gebrek aan informatie maakt het moeilijk om de potentiële risico's van stoffen in water in te schatten. Vaak kan op basis van het concept *"threshold of toxicological concern"* (TTC) voorzichtig een veilige grenswaarde voor de gezondheid worden afgeleid. In de praktijk is het echter doeltreffender om een nauwkeuriger beoordeling van potentiële risico's te maken op basis van een *in silico* benadering om onnodige maatregelen en bezorgdheid als gevolg van een conservatieve beoordeling te voorkomen.

Aanpak: combineren van resultaten van QSAR, read-across en bioassays

In het huidige BTO-project werden zestien waterrelevante stoffen geprioriteerd voor QSAR en read-across op bacteriële mutageniteit met behulp van *in silico* modellen, waaronder de QSAR Toolbox, VEGA QSAR en CASE Ultra.

QSAR
Als de identiteit of structuur van een stof in water bekend is, kunnen *in silico* modellen worden gebruikt om de toxiciteit te voorspellen. QSARs laten zien welke chemische substructuren en/of fysisch-chemische eigenschappen een bepaald type toxiciteit kunnen voorspellen. Op basis van de chemische structuur kan worden bepaald of een nieuwe stof een *"structural alert"* heeft voor een bepaald effect. Een *structural alert* betekent dat de stof op basis van de chemische structuur een bepaald effect kan vertonen, rekening houdend met het feit dat in de praktijk andere eigenschappen de mate van toxiciteit en de potentiële risico's beïnvloeden.

Read-across
Voor een meer betrouwbare voorspelling kan de stof door middel van een zogenaamde read-across worden vergeleken met stoffen met een soortgelijke chemische structuur en fysisch-chemische eigenschappen. Op die manier kunnen de aannemelijke schadelijke effecten van een stof op de menselijke gezondheid en het milieu worden beoordeeld.

Toxicologische eindpunten
Daarnaast is voor zeven geselecteerde stoffen (een selectie op basis van DNA-reactiviteit en oplosbaarheid in water) een Ames-fluctuatietest uitgevoerd om de bacteriële mutageniteit te onderzoeken. Ook werden zes stoffen geselecteerd voor QSAR en read across op oestrogene activiteit in de QSAR Toolbox en VEGA QSAR op basis van structurele waarschuwingen voor oestrogene activiteit.

Resultaten: *in silico* tools en bioassays in een geïntegreerde teststrategie
Voor alle geselecteerde verbindingen voor beoordeling van bacteriële mutageniteit werd bij combinatie met *in silico* verkregen resultaten een negatieve of onbesliste voorspelling gedaan. De Ames fluctuatietest bevestigde de negatieve voorspellingen voor de zeven beoordeelde stoffen. Bevestiging van positieve voorspellingen zou de waarde van de combinatie van *in silico* tools en bioassays sterker hebben gemaakt, maar dergelijke verbindingen ontbraken in de huidige dataset. Niettemin werden geïntegreerde teststrategieën ontwikkeld voor afzonderlijke stoffen en watermonsters, waarbij het in het laatste geval ging om complexe mengsels van verbindingen in lage concentraties in water. *In silico* tools kunnen niet worden gebruikt zonder chemische analyse om de stoffen te identificeren die van belang zijn bij de beoordeling van de watermonsters. Een *in silico* aanpak is sneller en kosteneffectiever voor individuele stoffen zonder toxiciteitsgegevens dan het uitvoeren van toxicologische studies. Bioassays kunnen worden gebruikt om *in silico* verkregen resultaten te verifiëren.

Toepassing: gecombineerd gebruik van in silico-tools en bioassays heeft toegevoegde waarde
In silico tools geven inzicht in welke microverontreinigingen in het water een risico voor de gezondheid en het milieu kunnen vormen. In combinatie met bioassays kunnen de verkregen resultaten worden geverifieerd. Daardoor hebben ze een duidelijke meerwaarde voor het oplossen van diverse waterkwaliteitsvraagstukken rond het prioriteren van vervolgonderzoek voor veiligheids- en risicobeoordeling en zorgen ze ervoor dat maatregelen en besluitvorming om deze risico's te beperken doelgericht kunnen worden gekozen en ingezet. Ontwikkelaars van *in silico* tools breiden voortdurend de databanken en algoritmen uit om de voorspellende capaciteit van hun instrumenten te optimaliseren, met name voor de meer complexe toxicologische eindpunten zoals reproductietoxiciteit, neurotoxiciteit en immunotoxiciteit.

Het Rapport
Dit onderzoek wordt gerapporteerd in Evaluation of QSAR tools in combination with bioassays for transformation products and emerging substances (BTO-2023.015).

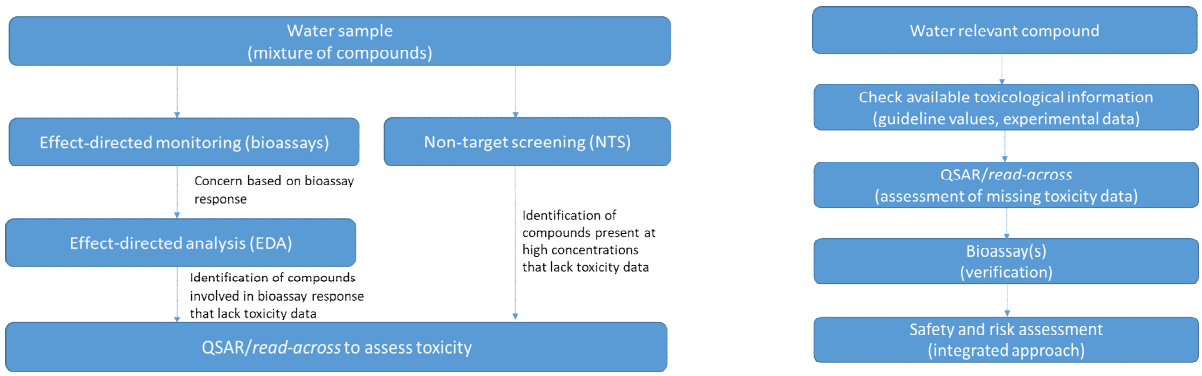
Andere relevante publicaties:
- Toepassen van QSAR en read-across modellen voor waterkwaliteit, H_2O online maart 2022
- Tools for human health risk assessment of emerging chemicals, BTO 2018.030.
- The Threshold of Toxicological Concern (TTC): refinement of the concept and application to drinking water, BTO 2016.069.

Management summary

Evaluation of QSAR tools in combination with bioassays for transformation products and emerging substances

Authors Astrid Reus, MSc., dr. Renske Hoondert, dr. Miina Yanagihara

The drinking water sector is confronted with many challenges in terms of water quality of their sources due to chemical threats. Due to a lack of toxicological information for many emerging substances, assessing any possible risks for public health and the environment is not always possible. *In silico* (computer) tools aimed at predicting the toxicity of substances offer a solution for this. Based on the chemical structure, the potential adverse effects of a substance can be assessed using quantitative structure-activity relationships (QSAR) or read-across to closely related substances. Our research shows that cell-based methods using bacteria or cells (bioassays) can be used to verify the results obtained with *in silico* tools. This information action perspectives for preventing risks to human health and the environment.



Integrated testing strategies for evaluating water samples (left) and individual water-relevant substances (right)

Importance: assessing potential risks of substances in water

Low concentrations of many different substances with varying physicochemical properties can be found in drinking water sources. For new, emerging substances, there is often little information available on toxicological properties and potentially adverse effects on human health and the environment. In an unusual situation or an emergency, when water suddenly contains a high concentration of a chemical, it is essential to quickly assess the potential risks to drinking water production, human health and the environment.

A lack of information makes it difficult to assess the potential risks of substances in water. It is often possible to conservatively derive a safe limit value for health based on the threshold of toxicological concern (TTC) concept. In practice, however, it is more effective to make a more accurate assessment of potential risks based using an *in silico* approach to avoid unnecessary measures and concerns due to a conservative assessment.

Method: combining results of QSAR, read-across and bioassays

In the current BTO project, sixteen water-relevant substances were prioritized for QSAR and read-across on bacterial mutagenicity using *in silico* models including th QSAR Toolbox, VEGA QSAR and CASE Ultra.

QSAR

If the identity or structure of a substance in water is known, *in silico* models can be used to predict the toxicity. QSARs show which chemical substructures and/or physicochemical properties could predict a particular type of toxicity. Based on the chemical structure, one can determine whether a new substance has a 'structural alert' for a specific effect. A structural alert means that based on the chemical

structure the substance may exhibit a particular effect, keeping in mind that, in practice, other properties affect the degree of toxicity and potential risks.

Read across
For a more reliable prediction, the substance can be compared by a so-called read-across with substances with a similar chemical structure and physicochemical properties. In this way, it is possible to assess the plausible adverse effects of a substance on human health and the environment.

Toxicological endpoints
In addition, an Ames fluctuation test was performed for seven selected substances (a selection based on DNA reactivity and water solubility) to investigate bacterial mutagenicity. Also, six substances were selected for QSAR and read across on estrogenic activity in the QSAR Toolbox and VEGA QSAR based on structural alerts for estrogenic activity.

Results: *in silico* tools and bioassays in an integrated testing strategy
All compounds selected for assessment of bacterial mutagenicity were predicted negative or inconclusive when combining the results obtained with the *in silico* tools. The Ames fluctuation test confirmed the negative predictions for the seven substances evaluated. Confirmation of positive predictions would have made the value of combining *in silico* tools and bioassays stronger, but such compounds were lacking in the current dataset. Nevertheless, integrated testing strategies were developed for individual substances and water samples, the latter being complex low-level mixtures of compounds in water. *In silico* tools cannot be used without chemical analysis to identify the substances of interest when evaluating the water samples. An *in silico* approach is faster and more cost-effective for individual substances without toxicity data than conducting toxicological studies. Bioassays can be used to verify the results obtained with *in silico* tools.

Implementation: combined use of *in silico* tools and bioassays has added value
In silico tools provide insight into which micro-pollutants in water may pose a risk to health and the environment. When combined with bioassays, the results obtained can be verified. As a result, they have a clear added value for solving various water quality questions around prioritizing follow-up research for safety and risk assessment and ensuring that measures and decision-making to mitigate these risks can be purposefully chosen and deployed. Developers of *in silico* tools are continuously expanding the databases and algorithms to optimize the predictive capacity of their tools, especially for the more complex toxicological endpoints such as reproductive toxicity, neurotoxicity and immunotoxicity.

The Report
This research is reported in *Evaluation of QSAR tools in combination with bioassays for transformation products and emerging substances* (BTO-2023.015).

Other relevant publications:
- Toepassen van QSAR en read-across modellen voor waterkwaliteit, H_2O online maart 2022
- Tools for human health risk assessment of emerging chemicals, BTO 2018.030.
- The Threshold of Toxicological Concern (TTC): refinement of the concept and application to drinking water, BTO 2016.069.

Contents

1 Introduction

Production and supply of safe drinking water is the primary task of drinking water companies. A particular challenge is related to the water quality of the raw water sources. Surface water as well as groundwater contains low concentrations of chemical substances with diverse physical-chemical properties (Houtman et al. 2013, Rozemeijer and Broers, 2007, Schipper et al. 2008, Ter Laak et al. 2012). For new, emerging substances, information on possible adverse effects to human health and the environment is often not available. In case of an emergency, when the water suddenly contains a chemical substance at a high concentration, rapid assessment of possible consequences for drinking water production and the environment is essential. Since water sources can be contaminated with many different chemical substances at low concentrations, various treatment processes can be needed to prepare safe drinking water, especially in the case of using surface water as a source (Di Marcantionio et al. 2020, Kegel et al. 2010, Verliefde et al. 2007). During water treatment, byproducts can be formed for which toxicological properties are usually unknown and it is acknowledged that these transformation products can be more toxic than the parent compounds (Brunner et al. 2019, Sharma et al. 2018). In summary, at different stages in drinking water production it is often difficult to assess potential risks for human health and the environment due to lack of toxicological information.

Effect-based monitoring using bioassays is a valuable approach to assess the potential of substances to cause adverse effects to human health and the environment. Bioassays are experiments with cells or bacteria (*in vitro*) or with living organisms, invertebrates such as water fleas and mussels (*in vivo*) and can be performed in laboratories or in the field (*in situ*) (Brack et al. 2016, Robitaille et al. 2022). Bioassays are cost-effective when a water sample is tested as a mixture of chemical substances and toxicity can be assessed without identifying substances by chemical analysis. It is possible to perform bioassays also for new, emerging substances and transformation products if the identity and structure is known and sufficient test material is available, however, depending on the number of different bioassays this can be time consuming and costly. Since many adverse effects are possible (for example, endocrine disruption, effects to the nervous system and DNA damage), a factorial large number of bioassays is needed for toxicity assessment of all water samples and all individual (emerging) compounds (Schriks et al. 2015).

In silico (computer) models focusing on toxicity assessment offer a possible solution (Ellison et al. 2010, Simon-Hettich et al. 2006). These models contain databases with information of different adverse effects for many substances. Quantitative Structure-Activity Relationship (QSAR) models can be used to indicate the presence of 'structural alerts' for a certain adverse effect. These structural alerts are substructures or functional groups within chemical structures that are often associated with adverse effects (Benigni 2004, Benigni and Bossa 2008, Benigni et al. 2013, Cronin et al. 2017). In addition, the chemical structure of the target compound can be compared to substances with chemical similarity (read-across) to estimate possible adverse effects to human health and the environment (Benfenati et al. 2019, Hewitt et al. 2010). For individual substances lacking toxicological information, the *in silico* approach can be less time-consuming and more cost-effective than conducting bioassays. A prerequisite of using *in silico* models is availability of the target compound chemical identity.

This project is a follow-up of the BTO report 2018.030 '*Tools for human health risk assessment of emerging chemicals*'. In this report a method for structured toxicological evaluation of chemicals has been described using existing data and in silico tools such as QSAR, read across and the TTC concept. The report also addressed the functional properties of various *in silico* tools. The current BTO project updates this knowledge, developed manuals for two *in silico* tools and presents additional research on combining *in silico* tools with bioassays.

The aim of this BTO project was to develop an effective and practical strategy for human risk assessment of substances lacking toxicological information that can be used to assess new, emerging substances in drinking water sources and in case of emergencies. The strategy focused on two toxicological endpoints relevant for drinking water and distinguishes between

1) prioritization of substances that are most relevant for human health (useful for risk assessment, purification effort, source tracking and lobby)
2) rapid assessment of hazard and risks in case of emergencies.

The project was focused on the implementation of the endpoint mutagenicity (a mechanism of DNA damage) and two different software applications, *i.e*. the OECD QSAR Toolbox (www.qsartoolbox.org) and VEGA QSAR (Benfenati et al. 2013, https://www.vegahub.eu/portfolio-item/vega-qsar/). Herewith, the added value of combining bioassays and *in silico* models was studied, resulting in availability of toxicological data of selected substances and information on the position of bioassays and *in silico* models in a strategy for hazard and risk assessment.

2 *In silico* tools for toxicological assessment

In silico (computational) toxicology refers to the prediction of the toxicity of a chemical from its molecular structure (QSAR) and assessment from the properties of similar compounds whose toxicity is known (read-across). It is rapid, economic and animal-free, and besides its applications in regulatory toxicology, it can also be applied to water relevant substances. *In silico* models contain databases with information of different adverse effects for many substances. A prerequisite of using *in silico* models is availability of the target compound chemical identity, as the CAS number, chemical name or structure is used as input for the tools. QSAR models can be used to indicate the presence of 'structural alerts' for a certain adverse effect. In addition, the chemical structure of the target compound can be compared to substances with chemical similarity (read across) to estimate possible adverse effects to human health and the environment. For individual substances lacking toxicological information, an *in silico* approach is less time consuming and thus more cost-effective than conducting bioassays (Benigni et al 2020, Yang et al. 1998, Wichard 2017).

The QSAR Toolbox is a free software application which is often used in the field of *in silico* toxicology. The QSAR Toolbox is funded and co-owned by the Europen Chemical Agency (ECHA) and the Organisation for Economic Cooperation and Development (OECD), where OECD is leading in the updates 1-2 times a year. The multiple varieties and functionalities resulted in the widely acceptance of this QSAR approach in a diverse range of governmental organizations, research institutions and industry (www.qsartoolbox.org). The toxicological endpoint of interest can be selected and predicted by internationally harmonized methods, such as the use of empirical data of analogues and the application of trend-analysis, read-across and available QSAR data. The functionalities that the program offers are predicting potential hazards by profiling the target molecule and its metabolites, identifying analogues of a target chemical, retrieving experimental results available for those analogues, and fill data gaps by trend-analysis, read-across, and QSARs. Prediction of toxicity by data gap filling involves manual steps (category definition and subcategorization) and expert judgment. The QSAR Toolbox consists of six modules, *i.e.* 1) Input, 2) Profiling, 3) Data, 4) Category Definition, 5) Data Gap Filling and 6) Report. The first three modules can be considered as the workflow set up; input of the target compound, definition of the query and which data should be included in the query. The last three modules consist of data gathering and prediction parts, to classify the target chemical(s) by grouping them with analogues based on chemical properties. All modules contribute to the gathering of endpoint specific data of the target chemical (Dimitrov et al. 2016). Standardized, automated workflows have at this point been developed for three endpoints, *i.e.* 1) ecotoxicity, 2) skin sensitization and 3) skin sensitization for defined approaches. A workflow can be described as selections and preferences from start (Input) to end (Report). It is very likely that the workflows will be extended to other endpoints in subsequent versions of the QSAR Toolbox (Yordanova et al. 2019).

Initially, the suitability of Generalized read-across (GenRA) was explored, which is a recently developed read-across tool of the US EPA (Shah et al. 2016) and was proposed as a simpler and more rapid alternative to the QSAR Toolbox for hazard and risk assessment of water relevant substances. GenRA is incorporated into the US EPA CompTox Chemicals Dashboard[1] and is able to predict responses related to *in vivo* toxicity (based on ToxRefDB[1,2] (Martin et al. 2009)). A preliminary study with the GenRA revealed that its functionalities did not match the need for water quality assessment, as the link to water-relevant endpoint, such as genotoxicity and endocrine disruption, was not clear (it links to ToxRefDB and ToxCast[1] (Dix et al. 2007)) and the analogue chemicals were often considered not relevant (no possibility to change manually). Based on a systematic comparison of other available software applications (Table 1), VEGA HUB was found to be the most interesting, which was therefore selected for further exploration instead of the GenRA.

The VEGA HUB is a freely available platform offering a wide collection of QSAR and read-across software and models. All software applications presented in the VEGA HUB[3] are in line with the REACH and the ICH M7 guidelines (ECHA 2008, EMA 2015). Similar to the QSAR Toolbox, the VEGA HUB continuously develops by feedback and comments from the users and evaluators. The evaluators consist of experts from companies, institutes, authorities, and regulators. VEGA QSAR is a comprehensive collection of QSAR models, constructed in well-known QSAR software tools such as CEASAR and SarPy (Benfenati et al. 2013). All the specified models contain target endpoints for regulatory purposes similar to the endpoints offered in the QSAR Toolbox, *i.e.* toxicology, ecotoxicology, environmental fate, and physical chemical properties. VEGA QSAR includes both expert rule-based and statistical-based models. By using ToxRead and ToxWeight it is possible to predict the biological activity of the target chemicals based on known values for structurally similar substances. The algorithms in these read-across approaches are independent of the VEGA QSAR model algorithms and can thus be considered as two independent prediction software programs. The algorithms used in the read-across approaches analyze and define the role of fragments and descriptors for both the target chemical and the analogues. The combination of these programs provides the user with a potential workflow in the same line as constructed in the QSAR Toolbox. Like the QSAR Toolbox, in ToxRead/Toxweight read-across can be performed for one chemical at a time and it is not possible to view the analogues, which disables expert judgement. Therefore, the current BTO project focused on the VEGA QSAR only.

Since VEGA QSAR has some limitations with respect to adjusting the selection of analogues (Table 1), the applicability of CASE Ultra[4] (Chakravarti et al. 2012) was also investigated. CASE Ultra is a commercially available software application for building predictive models and exploring the underlying mechanisms of biological activity of chemicals. It has both expert rule-based and statistical-based methodologies built in for a complete ICH M7 compliant assessment. CASE Ultra is user friendly and many of its models are fully supported by US FDA and other regulatory agencies around the world. CASE Ultra is quite similar to the QSAR Toolbox in terms of the general workflow, functionalities, and data gap filling methods. Similar to the VEGA QSAR, CASE Ultra offers expert rule-based and statistical-based models. CASE Ultra differs from the QSAR Toolbox in completeness as it haves automated workflows, more extensive graphs and output results, more guidance in the selections, and more user friendly interfaces. In addition, CASE Ultra includes US databases with experimental data on toxicological endpoints, which are complementary to the European data from QSAR Toolbox.

[1] https://www.epa.gov/
[2] ToxRefDB provides detailed chemical toxicity data (mammalian toxicity information) in a publically accessible searchable format
[3] www.vegahub.eu
[4] http://www.multicase.com/case-ultra

Many more *in silico* tools available on toxicity prediction, including the Toxicity Estimation Software Tool (TEST)[5], Ecological Structure Activity Relationships (ECOSAR) Predictive Model[5], EPI (Estimation Program Interface) Suite[5], Leadscope Model Applier[6], ToxTree[7], LAZAR[8], Spartan[9] and SPARC Performs Automated Reasoning in Chemistry[10]. In addition, QSAR models can be developed by institutes themselves (see Table 2). TEST, EPI Suite, ToxTree, QSAR Toolbox and VEGA have also been investigated previously, in which QSAR Toolbox was defined as the most complete tool as it had one of the largest collections of publicly available data and the most extensive range of analysis options of the available tools at that time (BTO 2018.030). It was also recognized that *in silico* tools are under development and that continuous inventarisation is warranted. QSAR Toolbox, VEGA and CaseUltra were selected for further research in the current project because of their notoriety in the field of *in silico* toxicity assessment and the possibility to perform QSAR and read-across in the same model. Nevertheless, the models mentioned here can be useful for chemical screening and prioritization for further study. Leadscope Model Applier was developed in close collaboration with the US Food and Drug Agency (FDA) and was thus primarily developed for pharmaceuticals (Roberts et al. 2000). The software is now commercialized via Instem[5] and it has been expanded with compounds other than pharmaceuticals. In tems of applicability for toxicity assessment it is assumed that Leadscope Model Applier is comparable to CASE Ultra, but this was not verified as the software was not freely accessible. A comparison of the functional properties of the QSAR Toolbox, VEGA and CASE Ultra is shown in Table 1.

Based on Table 1, the differences between the QSAR Toolbox, VEGA HUB and CASE Ultra are limited, the most prominent differences are highlighted below. The interface of the QSAR Toolbox is more complicated and the software application requires training before use, it is considered user friendly after acquiring training. With VEGA QSAR it is possible to predict toxicity of multiple chemicals and multiple endpoints simultaneously, whereas the QSAR Toolbox only enables individual prediction of a single endpoint on a single (target) chemical. Within CASE Ultra it is possible to run an indefinite large number of chemicals simultaneously, with a recommended maximum of 10,000 chemicals at a time. Although high throughput is possible in VEGA QSAR and CASE Ultra, it is not possible to adjust the selection of analogous chemicals that are considered to be the most structurally similar to the target compounds, a feature that is possible in QSAR Toolbox. Consequently, supported by the possibility to adjust descriptors, the predictions of the QSAR Toolbox may be more reliable than those of the VEGA QSAR and CASE Ultra. However, expert judgement from a chemical background point-of-view is always needed to provide argumentation for acceptance or rejection of a response in any case. Out of the three tools studied, only CASE Ultra provides some sort of uncertainty estimate to quantify model reliability.

The acceptance of predictions obtained with QSAR models in regulatory settings is dependent on individual regulatory systems in different OECD[11] member counties. The International Council for Harmonisation of technical requirements for pharmaceuticals for human use (ICH) M7(R1) guideline on 'Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk' describes the rule of thumb of combining at least one expert rule-based model and one statistical-based model to receive a convenient toxicity prediction with QSARs (EMA 2015). QSAR Toolbox is an expert-rule based model and should therefore be supported by at least one other software application. VEGA QSAR and CASE Ultra contain both expert rule-based and statistical-based models and are therefore complementary to the QSAR Toolbox. More background on the QSAR Toolbox, VEGA QSAR and CASE Ultra can be found in documentation of the models (www.qsartoolbox.org; Benfenati et al. 2013; https://www.vegahub.eu/portfolio-item/vega-qsar/; Chakravarti et al. 2012; http://www.multicase.com/case-ultra).

[5] https://www.epa.gov/
[6] www.instem.com
[7] https://toxtree.sourceforge.net/
[8] https://lazar.in-silico.ch/predict
[9] https://www.wavefun.com/
[10] http://www.archemcalc.com/sparc.html
[11] Organisation of Economic Co-Development (www.oecd.org)

Table 1: Systematic comparison of the functional properties of the QSAR Toolbox, VEGA HUB and CASE Ultra

	QSAR Toolbox	VEGA HUB (combination of VEGA QSAR and ToxRead)	CASE Ultra
Clear chemical input and clear definition of the toxicological endpoint?	YES	YES	YES
Adjustable descriptors[a]?	YES	NO	NO
Adjustable selection of analogues[b]	YES	NO	NO
Additional input of physicochemical parameters required?	NO	NO	NO
High throughput possible?	NO	YES	YES
Models pre-defined[c] in tool?	NO	YES	YES
Prediction of toxicity possible with QSAR, read across and or/trend analysis?	YES	YES (read across by using ToxRead)	YES
Access to databases?	YES (partly)	NO	YES (partly)
Option to save the output as a report?	YES	YES	YES
Quantitative output options?[d]	YES	YES	YES
User friendliness (expert judgement on a scale from 1 (low) to 5 (high))	4	4	4
Complexity (expert judgement on a scale from 1 (low) to 5 (high))	1	5	4
Free availability	YES	YES	NO
Programming features available for e.g. R, KNIME, Windows script, Python)	YES (direct KNIME plug ins)	NO (written in C#)	NO

[a] Descriptors are the chemical characteristics of a molecule in numerical form (Danishuddin and Khan, 2016)
[b] An analogue is a compound having a structure similar to that of another compound, but differing from it in respect to for example, one or more atoms, functional groups, or substructures, which are replaced with other atoms, groups, or substructures (https://en.wikipedia.org)
[c] Either a model can be selected or it is clear which model is being used for the prediction
[d] Not applicable to qualitative endpoints

Although most *in silico* tools are user friendly, to generate reliable output it is essential to follow a training course on the application of these methods. Examples of training courses on the OECD Toolbox specifically and *in silico* toxicology generally are those organized by the REACH monitor and the International QSAR foundation (official organizers of OECD QSAR Toolbox trainings) and the Fraunhofer ITEM respectively. The practical skills obtained with the QSAR Toolbox and VEGA QSAR during the BTO project are described in manuals (Supplementary Information I). The manuals provide user friendly and stepwise instructions on the different modules of the software applications. With the help of these manuals, one can learn to click on the right buttons to obtain reliable results. The manuals emphasize the need for expert judgement, but do not provide details on or further instructions for data interpretation. For that part, it is recommended to follow an *in silico* toxicology training course and refer to other experts for peer-to-peer review and discussions on interpretation. The interest of joining an expert group was inventoried. Discussions are ongoing if this can be organized within the Dutch toxicology community/water sector.

3 Literature search on the use of *in silico* tools combined with bioassays in (drinking) water quality practice

A targeted literature search was performed to search for recent papers on the application of *in silico* tools and bioassays for water quality assessment. The words 'QSAR', 'genotoxicity' (or 'mutagenicity'), 'bioassay' and 'water' were used as search terms. The search was performed on 27 January 2021. Several peer-reviewed papers were found to describe the use of *in silico* tools combined with bioassays in (drinking) water quality practice. These papers were summarized in Table 2.

Table 2: Summary of targeted literature search on the application of in silico models in combination with bioassays for water quality assessment (abbreviations are listed at the end of the table).

Author(s)	*In silico* model(s) used	Target chemical(s)	Bioassay(s)	Conclusion
de Barros et al. 2021	TEST	TPs of herbicide metribuzin	Estrogenic activity (YES), acute toxicity (*Artemia salina*), cytotoxicity (HepG2)	TPs are potentially more toxic than the precursor metribuzin
Carpinteiro et al. 2017	ECOSAR, TEST	TPs of pharmaceutical diazepam and related benzodiapines	Not performed	Some TPs could be more toxic/mutagenic than the precursor drug, confirmatory tests are needed
Chen et al. 2021	Developed by laboratory (University of Chinese Academy of Sciences, Bejing)	TPs of pharmaceutical levofloxacin	Genotoxicity: SOS/umu	The combination of bioassay, QSAR computation and chemical analysis would be an efficient method to screen toxic TPs under chlorination treatment
Cvetnic et al. 2019	Developed by laboratory (University of Zagreb)	36 aromatic pollutants (including aniline, phenol and toluene) and photooxidative intermediates	Acute toxicity (*V. fischeri)*	QSAR models have a significant potential in environmental risk assessment
Li et al. 2016	QSAR Toolbox	Various DBPs	Not performed	Chemical analysis (GC/MS) coupled to a QSAR model is a powerful and fast nontargeted screening technique for compounds for identification and prioritization of DBPs

Table 2: continued

Author(s)	*In silico* model(s) used	Target chemical(s)	Bioassay(s)	Conclusion
Lu et al. 2020	Developed by laboratory (Tongji University, Shanghai)	6 plant protection products (including (Glyphosate and Paraquat)	Not performed	QSAR models can be useful to derive concentration limits of chemicals outside the criteria of human drinking water
Mahmoud et al. 2014	CASE Ultra, OASIS[a], Leadscope	PhotoTPs of banned pharmaceutical thalidomide	Genotoxicity (Ames), acute toxicity (*V. fischeri)*	QSAR predictions were in contrast with Ames results, adverse effects of PTPs cannot be excluded, further research is warranted
Matsushita et al. 2018	ToxTree, VEGA (CAESAR), LAZAR, TEST	TPs of intermediate chemical 3-methyl-4-nitrophenol	Genotoxicity (Ames)	A combination of regression analysis[b], chemical analysis (MS/MS) and QSAR may be useful for prioritizing TPs for further study (e.g. Ames). Initial starting point was Ames on water samples
Matsushita et al. 2016	ToxTree, VEGA (CAESAR), LAZAR, TEST	TPs of contrast agent iopamidol	Genotoxicity (Ames)	A combination of chemical analysis (LC/MS) and QSAR may be useful for prioritizing TPs for further study (e.g. Ames). Initial starting point was Ames on water samples
Moon et al. 2020	EPI Suite™, ToxTree, LAZAR, VEGA (CAESAR)	PMTs and SVHCs (not specified)	Not performed	ToxTree showed highest sensitivity for carcinogenicity in terms of regulatory purpose, for mutagenicity ToxTree is effective in terms of specificity and accuracy compared to other QSAR models
Pérez-Garrido et al. 2008	Developed by laboratory (Catholic University of San Antonio, Central University of Las Villas, Vigo University)	DBPs (haloacetic acids)	Not performed	Predicted values were close to available experimental data
Qin et al. 2017	Developed by laboratory (Guilin University of Technology)	50 DBPs from 9 classes (including halomethanes, haloacetic acids and nitrosamines)	Interaction with DNA (*E.coli* +/- DNA), cytotoxicity *(V. fischeri),* interaction with proteins/peptides (*E. coli* +/- GSH)	QSAR models are useful for toxicity prediction of DBPs

[a] Model included in QSAR Toolbox

[b] Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome' or 'response' variable, or a 'label' in machine learning parlance) and one or more independent variables (often called 'predictors', 'covariates', 'explanatory variables' or 'features'). https://en.wikipedia.org

Table 2: continued

Author(s)	*In silico* model(s) used	Target chemical(s)	Bioassay(s)	Conclusion
Sanabria et al. 2021	VEGA, QSAR Toolbox, CASE Ultra	TPs of pharmaceutical anastrozole	Not performed	TPs were non-biodegradable and showed positive alerts for mutagenicity, some TPs required further study based on QSAR predictions on carcinogenicity and mutagenicity, performance of *in vitro* confirmatory tests is recommended
Shao et al. 2019	QSAR Toolbox	Chemicals identified with non-target screening of surface water	Genotoxicity (Ames, MN), developmental toxicity (Zebrafish embryo toxicity test)	*In silico* approaches were integrated with bioassays, literature data and chemical analysis to link genotoxic effects and hazardous compounds in surface water
Stalter et al. 2016	Spartan, SPARC, literature	50 DBPs from 9 classes (including halomethanes, haloacetic acids and nitrosamines)	Genotoxicity (umuC, Ames, p53-bla), oxidative stress (AREc32, ARE-bla), DNA transcription (NF-kb-bla), interaction with DNA (*E. coli* +/- DNA) cytotoxicity (*V. fischeri),* interaction with proteins/peptides (*E. coli* +/- GSH)	*In silico* tools provide a mechanistic understanding of results of effect-based measurements, indirect genotoxicity (e.g. via oxidative stress) is more plausible than direct DNA damage for most investigated DBPs
Wei et al. 2020	Developed by laboratory (Guanxi Medical University, Shenzen University, University of Illinois)	DBPs (haloacetonitriles)	Genotoxicity (comet assay), cytotoxicity (CHO-K1)	QSARs can detect toxicity of DBPs prior to effect-based methods
Ye et al. 2014	QSAR Toolbox	39 chemicals identified with non-target screening of surface water (including atrazine, nicotine, pyrene, phenol)	Genotoxicity (SOS/umu, MN)	SOS/umu and MN test are a useful tool for evaluation and classification of genotoxicity of complex mixtures, potential genotoxicants can be initially identified with additional information from chemical analysis and the QSAR toolbox
Zhang et al. 2020	Developed by laboratory	DBPs (halogenated aromatics)	Cytotoxicity (CHO-K1)	Major toxicity drivers among the target DBPs were identified, DBPs with the highest concentrations may not necessarily contribute to the highest proportions of overall toxicity.

Abbreviations: CHO: Chinese hamster ovary cells, DBP: disinfection byproduct, ECOSAR: ecological structure activity relationships (tool), EPI Suite: estimation program interface Suite™ (tool), GSH: glutathione, LAZAR: lazy structure activity relationships (tool), MN: micronucleus test, PMT: persistent, mobile and toxic substances, PTP: phototransformation product, SVHC: substances of very high concern, TEST: toxicity estimation software tool, TP: transformation product, YES: Yeast Estrogen Screen. Note: SOS/umu is not an abbreviation, but refers to the mechanism (SOS response/umuC gene) involved in the respective bioassay. The same holds true for AREc32, p53, Nrf2. HepG2 is a human hepatocellular carcinoma cell line. CAESAR, CASE Ultra, OASIS, SPARC and VEGA are tradenames of *in silico* tools and models.

From this targeted literature search the following can be concluded with regards to the state of science of the combined application of *in silico* models and bioassays:

- Target compounds studied for hazard and risk assessment were mainly disinfection byproducts (DBPs) or transformation products (TPs) (from pharmaceuticals and plant protection products);
- Various models were used, either free or commercially available tools, or self-build tools;
- *In silico* tools are often combined with bioassays and/or chemical analysis, in cases that:
    - QSARs tools were applied to prioritize the use of bioassays and/or chemical analysis;
    - Bioassays were applied to verify QSAR results;
    - The combination of bioassays and QSAR modelling was used to link genotoxic effects and hazardous compounds. responsible for the bioactivity.

4 Chemical and bioassay selection for experimental work

In the experimental part of the project, the combined application of *in silico* tools and bioassays was explored using individual substances (*i.e.* not water samples). Suggestions for water relevant substances lacking toxicological information were identified from previous KWR research (Baken et al. 2018, Brunner et al. 2019) and proposed by drinking water companies having surface water as source for their drinking water production. This resulted in a diverse list of chemical substances, including pharmaceuticals, industrial chemicals, pesticides, metabolites, fragrances, contrast media, drugs of abuse, compounds of natural origin and transformation products (Appendix I). A priori it was defined to include ten substances for the experimental work. For the substances, selection criteria on which the prioritization was based were 1) commercial availability and pricing, 2) DNA reactivity (obtained by profiling in de QSAR Toolbox) and 3) solubility in water and organic solvents. Information on commercial availability and pricing was collected at a leading supplier of chemicals (SigmaAldrich.com). Adverse effects on DNA (genotoxicity) was selected as primary toxicological endpoint to assess, both *in silico* and in bioassays because the involved cellular mechanisms are well understood (Basu and Nohmi, 2018; Chatterjee and Walker, 2017; Friedberg et al. 2004), the structure-activity relationships are well known (especially for mutagenicity) (Benigni and Bossa, 2008; Kazius et al. 2005; Plošnik et al. 2016), and a well-established *in vitro* testing strategy to evaluate the hazards of individual chemicals is available (ECHA, 2017; EFSA, 2011; EMA, 2008). In addition, it is known that potentially genotoxic DBPs and TPs can be generated during water treatment steps (Han et al. 2018, Mestankova et al. 2014). Moreover, DNA damage can eventually lead to tumour formation (Baan et al. 2019, Nohmi 2018), and thus can have a large impact on quality of life. DNA reactivity was determined with the QSAR Toolbox using the following profilers[12] related to carcinogenicity and genotoxicity, *i.e.* 1) Carcinogenicity (genotox and nongenotox alerts by ISS), 2) DNA alerts for AMES, CA and MNT by OASIS, 3) DNA binding by OASIS, 4) DNA binding by OECD, 5) Protein binding alerts for Chromosomal aberration by OASIS, 6) Toxic hazard classification by Cramer (extended), 7) *in vitro* mutagenicity (Ames test) alerts by ISS and 8) *in vivo* mutagenicity (Micronucleus) alerts by ISS. Data on solubility in water and organic solvents were obtained from PubChem. For most substances water solubility data were available, whereas information on solubility in organic solvents was not readily available. Based on the combined selection criteria, substances for further research were prioritized (Table 3). The complete chemicals list including prioritization steps is available as Supplemental Information.

Ten compounds were evaluated for their genotoxic (mutagenic) potential. From the available bioassays (Ames fluctuation test, UMU test, p53 CALUX, comet assay and micronucleus test), the Ames fluctuation test was selected for the experimental work, because of its common use for water quality assessment (including existence of an ISO standard) (ISO 11350:2012, Reifferscheid et al. 2012), its functional analogy to the classical Ames test, which is part of the regulatory testing strategy for pharmaceuticals, chemicals and food and feed ingredients (ECHA, 2017; EFSA, 2011; EMA, 2008) and its capability to indicate mutagenicity, which may produce heritable effects (Phillips and Arlt, 2009) for which the structure-activity relationships are well described (Benigni and Bossa, 2008; Kazius et al. 2005; Plošnik et al. 2016). The Ames fluctuation test is a more rapid version of the classical Ames test and requires less sample volume (Reifferscheid et al. 2012), which is critical for (extracted) environmental (water) samples.

.

[12] In the QSAR Toolbox, the "Profiling" module contains all the knowledge in the system coded in profiling schemes (profilers). The profilers identify the affiliation of the target chemical(s) to preliminary defined categories (functional groups/alerts).The outcome of the profiling determines the most appropriate way to search for analogues, but they are also useful for preliminary screening or prioritization of substances (www.qsartoolbox.org).

Table 3: Selected chemicals for mutagenicity assessment

Compound	CAS no.	Origin	Ames fluctuation test performed in current project?
Nicotine	54-11-5	Natural	Yes
Trifluoroacetic acid	76-05-1	Industrial	Yes
Methocarbamol	532-03-6	Pharmaceutical	Yes
Cotinine	486-56-6	Endogenous (mammalian) metabolite of nicotine	Yes
Hydroxyatrazine[13]	2163-68-0	Environmental (microbial) metabolite of herbicide atrazine	Yes
EDTA[4]	60-00-4	Industrial	Yes
Dinoterb	1420-07-1	Plant protection product	Yes
Acridone	578-95-0	Endogenous (mammalian) metabolite of pharmaceutical carbamazepine	Yes
Levonorgestrel	797-63-7	Pharmaceutical	Yes
Cyanopropanal[14]	3515-93-3	Industrial	Yes
Sulpiride	15676-16-1	Pharmaceutical	No
Levocetirizine	130018-77-8	Pharmaceutical	No
Hordenine	539-15-1	Natural	No
Benzamide, 2-amino-N-(1-methylethyl)-	30391-89-0	Environmental (microbial) metabolite of herbicide bentazon	No
Clomitrazole	23593-75-1	Antimycotic	No

The maximum test concentration of each substance to be tested in the Ames fluctuation test was based on reported concentrations of *in vitro* (genotoxicity) studies as found in scientific literature (PubMed) with a maximum of 10 micromolair (mM), if not limited by solubility in water and/or dimethyl sulfoxide (DMSO). DMSO is a common solvent for *in vitro* bioassays. Water solubility was considered because of the aqueous nature of the exposure medium of *in vitro* bioassays

In addition to mutagenicity, a pilot exercise was performed on water-relevant endpoints other than genotoxicity, including estrogenic activity, reproductive toxicity and developmental toxicity. These endpoints are considered relevant in water quality assessment because effects can occur after chronic exposure to low concentrations (Dingemans et al. 2019). In this pilot exercise, the chemicals of the list (Supplementary Information I) that passed selection criterion 1 (commercial availability and pricing) were evaluated using the QSAR Toolbox and the following selected profilers relevant for reproductive and developmental effects*:* 1) Estrogen Receptor Binding, 2) rtER Expert System, 3) Protein binding by OASIS, 4) Protein binding by OECD, 5) Protein binding potency GSH en 6) DART (Developmental and Reproductive Toxicology) scheme.

Subsequently, the QSAR Toolbox was used to determine whether data were already available with respect to endocrine disruption (in particular estrogenic activity) and *in vivo* developmental and reproductive toxicity. Compounds for which no data were found were selected for read-across and are summarized in Tables 4 and 5. Some of them overlapped with the selected compounds for mutagenicity assessment (Table 3).

[13] Eventually not included in the experiments because of problems with solubility
[14] Eventually not included in the experiments because the wrong substance was received

Table 4: Selected compounds for read across on estrogenic activity

Compound	CAS nr.	Origin
Levocetirizine	130018-77-8	Pharmaceutical
Hordenine	539-15-1	Natural
Benzamide, 2-amino-N-(1-methylethyl)-	30391-89-0	Environmental (microbial) metabolite of herbicide bentazon
Clotrimazole	23593-75-1	Antimycotic
1H-Indene, 2,3-dihydro-	496-11-7	Industrial
2-aminobenzoic acid	118-92-3	Industrial

Table 5: Selected compounds for read across on developmental and reproductive effects

Compound	CAS nr.	Origin
EDTA	60-00-4	Industrial
Metolachlor ESA	171118-09-5	Environmental TP of herbicide metolachlor
Metolachlor OA	152019-73-3	Environmental TP of herbicide metolachlor
Methane, bromotrichloro-	75-62-7	Industrial
Nicotine	54-11-5	Natural
Dinoterb	1420-07-1	Plant protection product
1H-Benzotriazole, 5-methyl-	136-85-6	Industrial
Diglyme	111-96-6	Industrial
Triglyme	112-49-2	Industrial

5 Methods

5.1 Ames fluctuation test

Chemicals were dissolved in DMSO one day before use in the first Ames experiment. Three concentrations of each chemical were tested, *i.e.,* the top dose and two dilutions with 2-fold spacing. Dose solutions were stored at <-18°C for later use. In the second Ames experiment, three concentrations of each chemical were tested, *i.e.,* an 8-fold dilution, 16-fold dilution and 32-fold dilution of the top dose. Actual concentrations shown in Tables 4 and 5 were calculated based on the actual amount of the chemicals weighed and the volume of DMSO added.

The Ames fluctuation test was performed according to Heringa et al (2011) with minor modifications with regards to cytotoxicity measurement and data interpretation. In brief, *Salmonella typhymurium* strains TA98 and TA100 (Xenometrix, Switzerland) were briefly exposed to the chemicals for 90 minutes in absence and presence of an exogenous metabolic activation system (rat liver S9 mix), resulting in four different test conditions (TA98-S9, TA98+S9, TA100-S9 and TA100+S9). The final concentration of DMSO in the culture medium was 2%. Negative controls (Evian mineral water extract), solvent controls (DMSO) and appropriate positive controls were run in parallel. Two independent experiments were performed, and all experiments were carried out in triplicate cultures. Cytotoxicity was measured in parallel in strain TA98 by measuring optical density of the bacterial culture shortly before and immediately after treatment according to ISO 11350 (2012). Bacterial density of each chemical was normalized to the solvent control and for evaluation of the genotoxic responses, ≥50% cytotoxicity was considered severe and 30-50% cytotoxicity was considered moderate. The number of yellow wells per 48 wells of each replicate culture were counted manually as a measure of mutagenicity. A response was considered positive for genotoxicity if the response of the sample was different from the negative control with a certainty of 99%, based on a binominal distribution (Heringa et al., 2011). A chemical was considered positive for mutagenicity if at least one of the test conditions showed a positive response. In addition to numerical significance, biological relevance (e.g., comparison with concurrent negative control and/or historical control data) was taken into account.

5.2 QSAR and read-across

The selected compounds were evaluated for bacterial mutagenicity and estrogen receptor activity in both the QSAR Toolbox and VEGA QSAR. In the QSAR Toolbox, the profiling was done in the same was as done for the chemical selection (Chapter 4). The next step was to gather toxicologically relevant data available in the QSAR Toolbox for each compound individually. For seven out of the ten compounds that were selected for testing in the Ames fluctuation test, data on bacterial mutagenicity appeared to be available within the QSAR Toolbox. Initially, it was assumed that all compounds of the list were data-poor, but during the analysis it appeared that the QSAR Toolbox database contains data that are not easily findable or accessible in any other way. It was concluded that data gathering using the QSAR Toolbox is essential before selecting compounds for or continuing with read-across. This step was therefore implemented in the selection of six additional compounds for mutagenicity (Table 3) and in the selection of compounds for estrogen receptor binding activity (Table 4) and developmental and reproductive effects (Table 5).

After profiling and data gathering, chemicals were grouped by category definition[15] and subcategorization[16] to identify a set of structurally relevant analogues (mechanistically and structurally similar chemicals) for read-across. This categorization step starts with broad, endpoint non-specific grouping based on chemical structure, followed by endpoint specific subcategorization to eliminate dissimilar chemicals (Figure 1) until only representative analogues remain. On the one hand the analogues need to be clearly relevant based on functional groups, on the other hand it needs to be avoided to end up with too few datapoints for filling the data gaps of the target chemical. The identified analogues were used to predict the mutagenicity and estrogen receptor binding activity of the selected compounds. The workflow that was followed for genotoxicity assessment in the QSAR Toolbox is shown in Figure 2, in which profilers and databases specific for mutagenicity are marked with an asterisk (*). The workflow that was followed for estrogen receptor binding activity is shown in Figure 3.

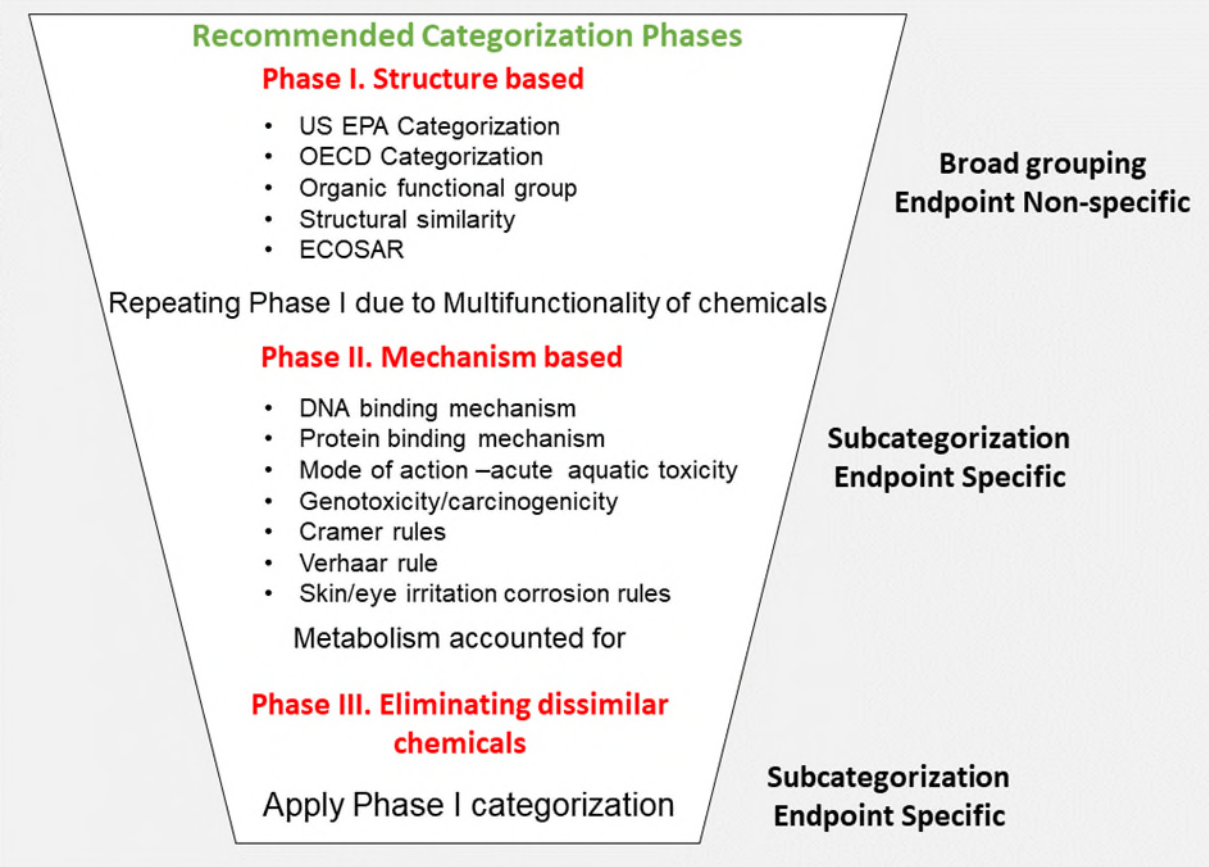


Figure 1: Recommended categorization phases in grouping of chemical for read across. Retrieved from: https://qsartoolbox.org/features/grouping/

[15] Category definition module provides the user with several means of grouping chemicals into a toxicologically meaningful category based on the specifics of the target molecule. The chemicals could be grouped according to different measures of "similarity" (structural or mechanistic similarity) so that, within a category data, gaps can be filled by read-across or trend analysis. This is the critical step in the workflow and several options are available in the Toolbox to assist the user in better definition of the category (www.qsartoolbox.org).

[16] Refinement of the category by removing the chemicals which differ mechanistically and/or structurally to the target chemical. (www.qsartoolbox.org).

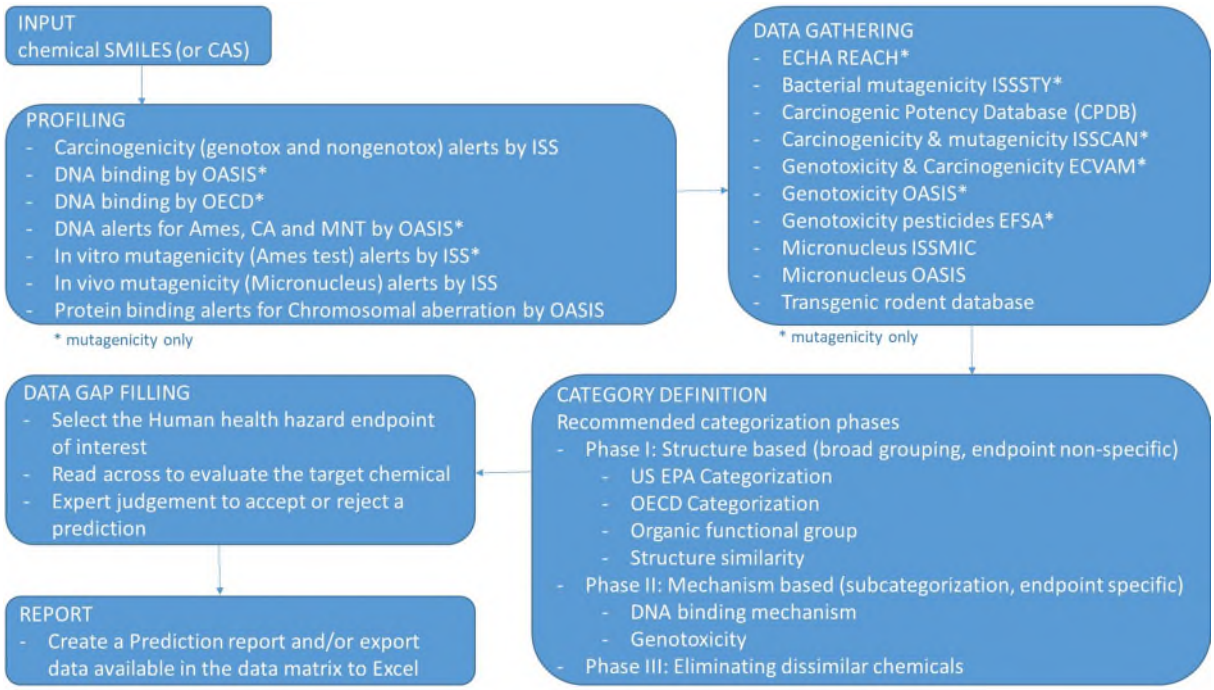


Figure 2: Workflow for assessment of genotoxicity in QSAR Toolbox. Profilers and databases specific for mutagenicity are marked with an asterisk (*).

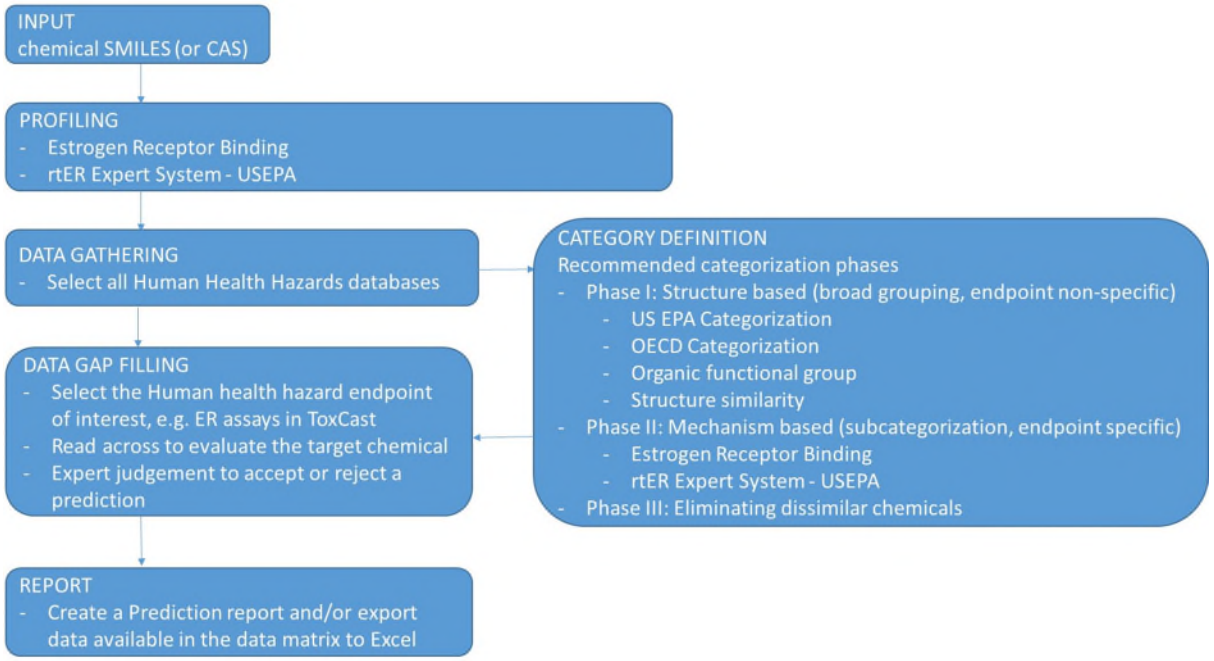


Figure 3: Workflow for assessment of estrogen receptor binding activity in QSAR Toolbox.

Following the prediction in the data gap filling phase[17], analogues which showed a dissimilar response from the predicted response of the target compound were discussed between toxicology and chemistry experts for their relevance in terms of structural and mechanistic similarity (expert judgement) to support or reject the prediction of the software. Inclusion/exclusion of the applicability domain[18] or parametric domain[19] was taken into account, but was not leading in the acceptance or rejection of the predicted result.

For mutagenicity assessment using the VEGA QSAR, the target chemicals were individually loaded into the software, followed by selection of four models relevant for bacterial mutagenicity, *i.e.,* 1) Mutagenicity (Ames test) model (CAESAR), 2) Mutagenicity (Ames test) model (SarPy/IRFMN), 3) Mutagenicity (Ames test)model (ISS) and Mutagenicity (Ames test) model (KNN/Read-Across). The workflow that was followed for genotoxicity assessment in the VEGA QSAR is shown in Figure 4, in which models specific for mutagenicity are marked with an asterisk (*).

For estrogen receptor binding activity, two relevant models were selected, *i.e.,* 1) Estrogen Receptor Relative Binding Affinity (IRFMN) and 2) Estrogen Receptor-mediated effect (IRFMN/CERAPP).

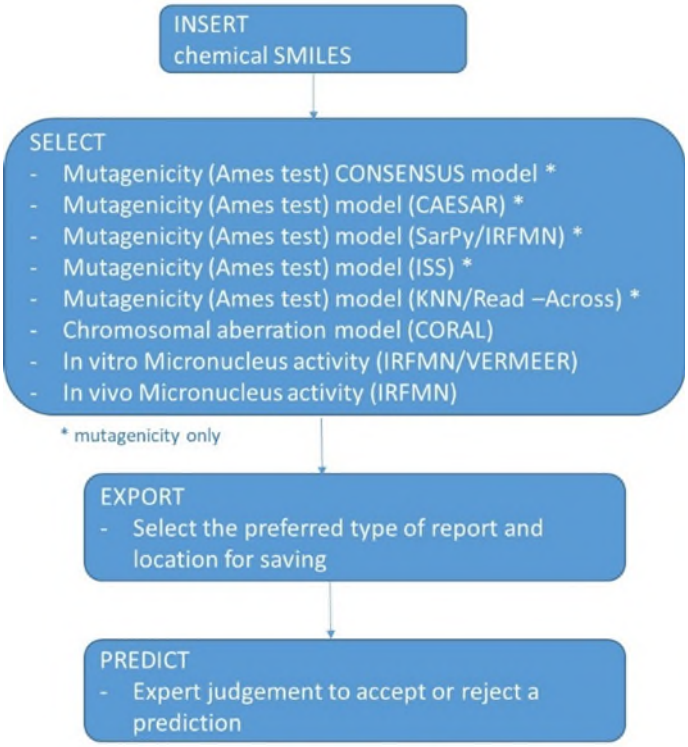


Figure 4: Workflow for assessment of genotoxicity in VEGA QSAR. Models specific for mutagenicity are marked with an asterisk (*).

[17] In the Data Gap Filling module the user is able to fill a data gap for their target substance using data from analogues with a trend analysis, read-across or existing QSAR models (www.qsartoolbox.org).
[18] The applicability domain (for both chemistry and machine learning) of a QSAR model is the physico-chemical, structural or biological space, knowledge or information on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds (en.wikipedia.org).
[19] The parametric domain is the space of possible parameter values that define a particular mathematical model (en.wikipedia.org)

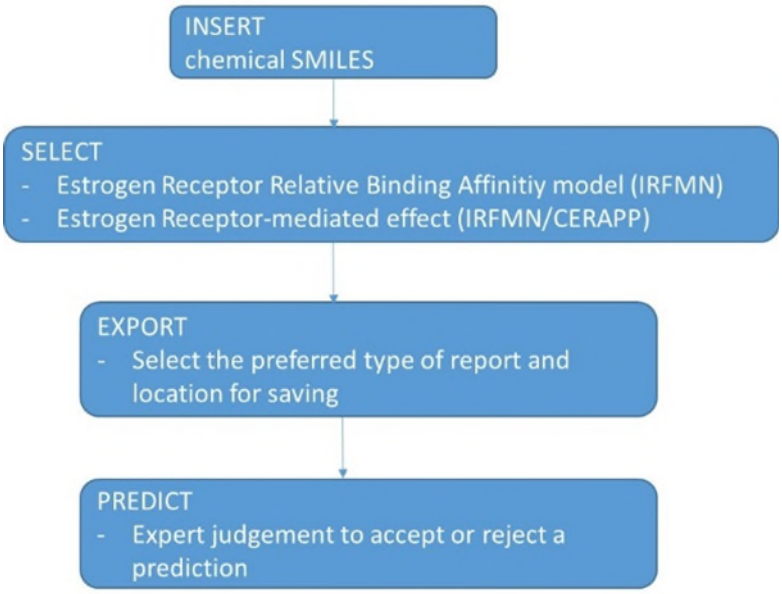


Figure 5: Workflow for assessment of estrogen receptor binding activity in VEGA QSAR.

In addition to QSAR Toolbox and VEGA QSAR, the sixteen compounds were investigated for mutagenicity using CASE Ultra. For mutagenicity assessment using CASE Ultra, the target chemicals were loaded into the software as a batch using the SMILES codes, following by selection of three models relevant for bacterial mutagenicity, *i.e.,* 1) an expert rule-based model: GT Expert Bacterial mutagenicity model, 2) a statistical-based model: GT1 BMUT OECD 471 Bacterial mutagenicity model, and 3) a statistical-based model: PHARM BMUT Statistical OECD 471 Bacterial mutagenicity model. The workflow that was followed for genotoxicity assessment in CASE Ultra is shown in Figure 6. The reliability (probability) as indicated by the software, was taken into account, but was not leading in the acceptance or rejection of the predicted result.

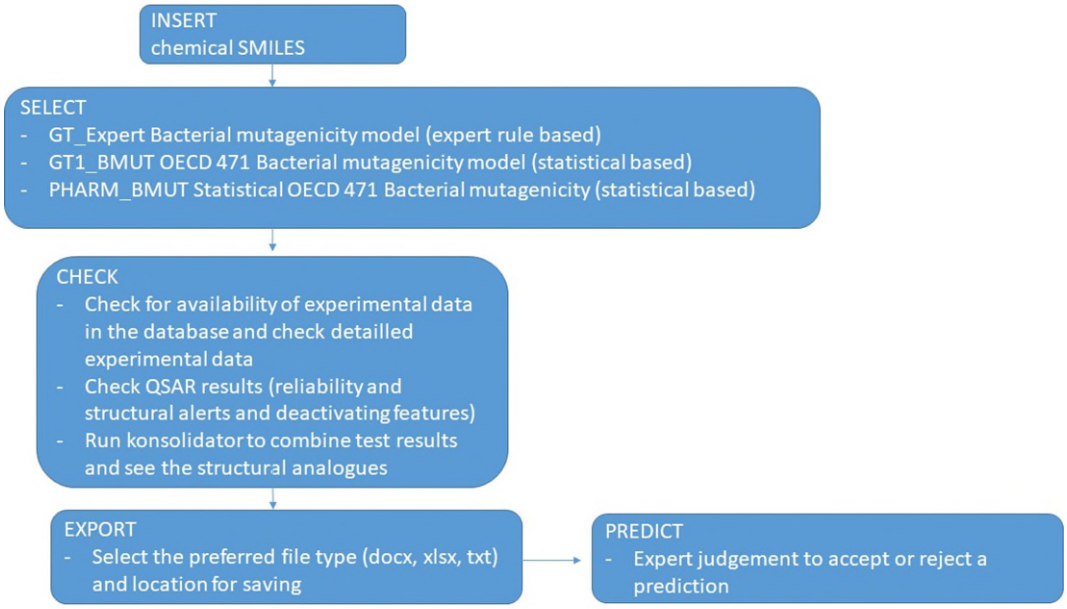


Figure 6: Workflow for assessment of mutagenic activity in CASE Ultra.

6 Results

6.1 Ames fluctuation test

Tables 6 and 7 show the experimental results of the first and second Ames experiment with the selected seven water-relevant compounds tested at the three concentration levels, respectively.

Table 6: Results of the first Ames experiment. Severe cytotoxicity (cell toxicity based on bacterial density) is highlighted in red and moderate cytotoxicity in orange. Negative cytotoxicity ≥20% was also highlighted in orange due to its relatively large deviation from 100%.

		Mean revertants/48 wells				Cytotoxicity (%)	
	Dose (mM)	TA98-S9	TA98+S9	TA100-S9	TA100+S9	TA98-S9	TA98+S9
Negative control		2	1	6	6	15	15
Solvent control		3	2	5	6	0	0
Positive control 1		44	48	45	39	54	0
Positive control 2		46				50	
Nicotine	10	2	1	5	5	61	-1
Nicotine	5	2	1	7	6	51	8
Nicotine	2.5	3	2	6	5	31	-23
Trifluoroacetic acid	10	3	2	7	5	29	-9
Trifluoroacetic acid	5	3	2	6	6	39	-12
Trifluoroacetic acid	2.5	2	1	6	7	35	7
Methocarbamol	10	2	1	5	5	47	3
Methocarbamol	5	3	2	7	6	39	0
Methocarbamol	2.5	2	3	5	5	32	-4
Cotinine	10	3	2	5	5	19	0
Cotinine	5	3	3	7	5	23	-20
Cotinine	2.5	1	1	3	5	35	3
Dinoterb	0.42	3	1	5	6	48	46
Dinoterb	0.21	3	2	5	5	37	35
Dinoterb	0.10	1	1	6	7	36	21
Acridone	0.53	3	2	5	6	36	-20
Acridone	0.26	3	1	4	7	53	-5
Acridone	0.13	1	1	3	4	36	23
Levonorgestrel	0.54	2	1	4	8	-19	3
Levonorgestrel	0.27	2	2	4	6	38	11
Levonorgestrel	0.13	1	2	5	7	23	-2

[a] According to ISO 11350:2012, cytotoxicity is measured in bacterian strain TA98 S9 only (both in the presence and absence of S9) due to the lower sensitivity of TA100.

In the first Ames experiment, all compounds showed moderate to severe cytotoxicity at one or more tested concentrations in strain TA98-S9 and to a lesser extent in strain TA98+S9. Consequently, a false negative response of the chemicals showing moderate or severe cytotoxicity cannot be excluded. Therefore, a second experiment was performed using lower concentrations of the same test compounds.

Table 7: Results of the second Ames experiment. Severe cytotoxicity is highlighted in red, a statistically significant positive mutagenic response is highlighted in orange.

		Mean revertants/48 wells				Cytotoxicity (%)	
	Dose (mM)	TA98-S9	TA98+S9	TA100-S9	TA100+S9	TA98-S9	TA98+S9
Negative control							
Solvent control		2	2	7	6	0	0
Positive control 1		39	42	46	47	57	15
Positive control 2		45				13	
Nicotine	1.25	2	1	8	7	16	14
Nicotine	0.625	0	2	8	5	12	17
Nicotine	0.3125	2	2	9	8	3	0
Trifluoroacetic acid	1.25	1	1	9	8	4	6
Trifluoroacetic acid	0.625	1	2	7	7	18	0
Trifluoroacetic acid	0.3125	2	1	10	7	22	18
Methocarbamol	1.25	2	1	8	7	12	7
Methocarbamol	0.625	1	0	9	6	17	12
Methocarbamol	0.3125	2	1	9	8	3	19
Cotinine	1.25	1	1	10	8	-11	12
Cotinine	0.625	2	1	9	8	16	-14
Cotinine	0.3125	1	1	7	4	16	-12
Dinoterb	0.0525	1	1	10	7	27	3
Dinoterb	0.02625	1	0	8	9	22	-3
Dinoterb	0.013125	3	1	8	7	7	-4
Acridone	0.06625	2	1	8	6	15	0
Acridone	0.033125	1	1	7	7	20	-7
Acridone	0.0165625	0	2	9	6	18	6
Levonorgestrel	0.0675	1	1	7	6	23	-18
Levonorgestrel	0.03375	2	2	11	7	20	0
Levonorgestrel	0.016875	1	1	7	8	0	4

[a] According to ISO 11350:2012, cytotoxicity is measured in bacterian strain TA98 S9 only (both in the presence and absence of S9) due to the lower sensitivity of TA100.

None of the test compounds showed severe or moderate cytotoxicity. The mid dose of Levonorgestrel was the only that showed a statistically significant genotoxic response. However, since the observed increase in mean number of revertants was not dose-related, this response may be considered not biologically relevant.

Based on the results of the two experiments and under the conditions used, all compounds were considered negative for mutagenicity in the Ames fluctuation test, except for Levonorgestrel which it is recommended to investigate if the observed response is reproducible.

6.2 QSAR and read across

6.2.1 Bacterial mutagenicity

In total sixteen water-relevant compounds were assessed for bacterial mutagenicity in the QSAR Toolbox and VEGA QSAR by three experts for internal review to strengthen the conclusion. In addition, the compounds were assessed for bacterial mutagenicity in CASE Ultra by one expert. Table 8 shows the summarized results. Since it is recommended to use two different models for *in silico* assessment, *i.e.* expert rule-based and statistical (Chapter 2), the conclusion was always based on two software applications, either QSAR Toolbox and VEGA QSAR or QSAR Toolbox and CASE Ultra.

Table 8: Summarized results of bacterial mutagenicity assessment of sixteen water relevant compounds using the QSAR Toolbox and VEGA QSAR.

Compound	QSAR Toolbox	VEGA QSAR	CASE Ultra	Combined conclusion QSAR Toolbox/ VEGA QSAR	Combined conclusion QSAR Toolbox/ CASE Ultra
Nicotine	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic
Trifluoroacetic acid	Non-mutagenic[a]	Likely non-mutagenic	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic
Methocarbamol	Likely non-mutagenic	Inconclusive	Likely non-mutagenic	Inconclusive	Likely non-mutagenic
Cotinine	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic	Non-mutagenic[a]	Non-mutagenic[a]
Hydroxyatrazine	Non-mutagenic[a]	Non-mutagenic	Likely mutagenic	Non-mutagenic[a]	Non-mutagenic[a]
EDTA	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic[a]
Dinoterb	Inconclusive	Inconclusive	Non-mutagenic	Inconclusive	Inconclusive
Acridone	Non-mutagenic	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic[a]
Levonorgestrel	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic[a]	Non-mutagenic[a]
Cyanopropanal	Non-mutagenic	Non-mutagenic	Likely non-mutagenic	Non-mutagenic	Likely non-mutagenic
Sulpiride	Likely non-mutagenic	Likely non-mutagenic	Likely non-mutagenic	Likely non-mutagenic	Likely non-mutagenic
Levocetirizine	Likely non-mutagenic	Inconclusive	Non-mutagenic[a]	Inconclusive	Non-mutagenic[a]
Trospium chloride	Inconclusive	Likely non-mutagenic	Non-mutagenic[a]	Inconclusive	Non-mutagenic[a]
Hordenine	Likely non-mutagenic	Inconclusive	Likely non-mutagenic	Inconclusive	Likely non-mutagenic
Benzamide, 2-amino-N-(1-methylethyl)-	Likely non-mutagenic	Inconclusive	Likely non-mutagenic	Inconclusive	Likely non-mutagenic
Clotrimazole	Likely non-mutagenic	Inconclusive	Non-mutagenic[a]	Inconclusive	Non-mutagenic[a]

[a] Based on experimental data

For eleven out of sixteen compounds, experimental data on bacterial mutagenicity was available within at least one software tool. Category definition and data gap filling was discontinued for compounds for which data on bacterial mutagenicity was available in the QSAR Toolbox.

6.2.2 Estrogenic activity

In total six compounds were assessed for estrogenic activity in the QSAR Toolbox and VEGA QSAR by one expert. Table 9 shows the summarized results. For reference, the results of the well-known estrogenic active compound estradiol are also presented.

Table 9: Summarized results of estrogenic activity of six water relevant compounds using the QSAR Toolbox and VEGA QSAR

Compound	QSAR Toolbox (AC50[a] in mg/L)	VEGA QSAR
Levocetirizine	Not predicted	Likely inactive
Hordenine	1-10 [b]	Likely inactive
Benzamide, 2-amino-N-(1-methylethyl)-	Not predicted	Likely inactive
Clotrimazaole	1-20[b]	Inactive
1H-Indene, 2,3-dihydro-	Not predicted	Likely inactive
2-aminobenzoic acid	1-5[b]	Likely inactive
Reference: Estradiol	0.0001-0.025[b,c]	Active[d]

[a] Active concentration 50, *i.e.* the concentration which gives 50% activation of the reporter gene studied.
[b] A range is presented based on the predicted results of different estrogen receptor bioassays
[c] Experimental data
[d] Conclusion on experimental result and model prediction was the same

With the QSAR Toolbox is it possible to predict bioassay responses in ToxCast bioassays. For which bioassays responses can be predicted depends on the availability of data of activity for the analogue substances. Different reporter gene assays[20] showed different quantitative results for the same endpoint (estrogen receptor activity), but AC_{50} values were in the same range. The variability in quantitative results can be explained by de different assay principles and mechanisms, further investigation was outside of the scope of the current project. For some chemicals there were no analogues with data on ToxCast bioassays relevant for estrogenic activity, resulting in no prediction. The VEGA QSAR provides qualitative responses on estrogen activity and predictions could be obtained for all chemicals studied.

[20] Attagene factorial cis ERE, NCGC Reporter Gene Assay ERa Agonist, Tox21_Era_BLA_Agonist_ch1, Tox21_Era_BLA_Agonist_ch2, Novascreen Human ER, and multiple OT_ER assays.

6.3 Combined results of Ames fluctuation test and read across

For seven out of sixteen compounds, the Ames fluctuation test was performed prior to read-across. The results were combined with those obtained with the *in silico* tools and are presented in Table 10.

Table 10: Combined results of the Ames fluctuation test and in silico models

Compound	QSAR Toolbox/VEGA QSAR	QSAR Toolbox/CASE Ultra	Ames fluctuationtest
Nicotine	Non-mutagenic[21]	Non-mutagenic[7]	Non-mutagenic
Trifluoroacetic acid	Non-mutagenic[7]	Non-mutagenic[7]	Non-mutagenic
Methocarbamol	Inconclusive	Likely non-mutagenic	Non-mutagenic
Cotinine	Non-mutagenic[7]	Non-mutagenic[7]	Non-mutagenic
Dinoterb	Inconclusive	Inconclusive	Non-mutagenic
Acridone	Non-mutagenic[7]	Non-mutagenic[7]	Non-mutagenic
Levonorgestrel	Non-mutagenic[7]	Non-mutagenic[7]	Likely non-mutagenic

[7] Based on experimental data

The results of the Ames fluctuation test showed good concordance with the *in silico* predictions and where the *in silico* predictions were inconclusive, a conclusion on mutagenicity could be drawn.

[21] Based on experimental data

7 Discussion and conclusion

7.1 Application of *in silico* predictions for water quality assessment

During this project, three important functionalities of *in silico* tools for water quality assessment were identified, *i.e.:*

1) prioritization of research needs based on structural alerts
2) read across for hazard and risk assessment;
3) discovery of hard-to-find experimental data on effects of substances.

It should be noted, however, that for hazard and risk assessment of individual substances, the use of experimental data (if available) should always be considered prior to *in silico* predictions. Moreover, experimental data will prevail over estimated results (ECHA 2016). If no or insufficient experimental data is available, *in silico* tools can be used for data gap filling. Conclusions based on estimated values should always be interpreted with caution, as it remains a prediction. When *in silico* tools reveal experimental data that was initially not found in scientific literature or databases, the reliability of the data should be assessed. Different starting points and strategies for the use of *in silico* tools, including chemical analysis and bioassays, are summarized in Figures 7 and 8, and discussed below.

Chemical analyses (targeted or non-targeted) are often used to assess chemical water quality (Been et al. 2021, Brunner et al. 2020). In addition, effect-based monitoring (bioassays) can be performed to examine biological effects of substances in water on living tissues and organisms. This can be done using bacteria and cell lines (*in vitro*) or living organisms such as water fleas and algae (*in vivo*). Important applications of *in silico* models for water quality monitoring are prioritization of substances for research and support in risk analysis of bioassay responses and chemical analyses (Figure 7) (Reus et al. 2022). Here, prioritization includes the selection of chemicals and/or locations for water quality monitoring using either (non-)targeted chemical analysis and/or bioassays. Furthermore, *in silico* data can also be used to substantiate decisions on monitoring frequencies.

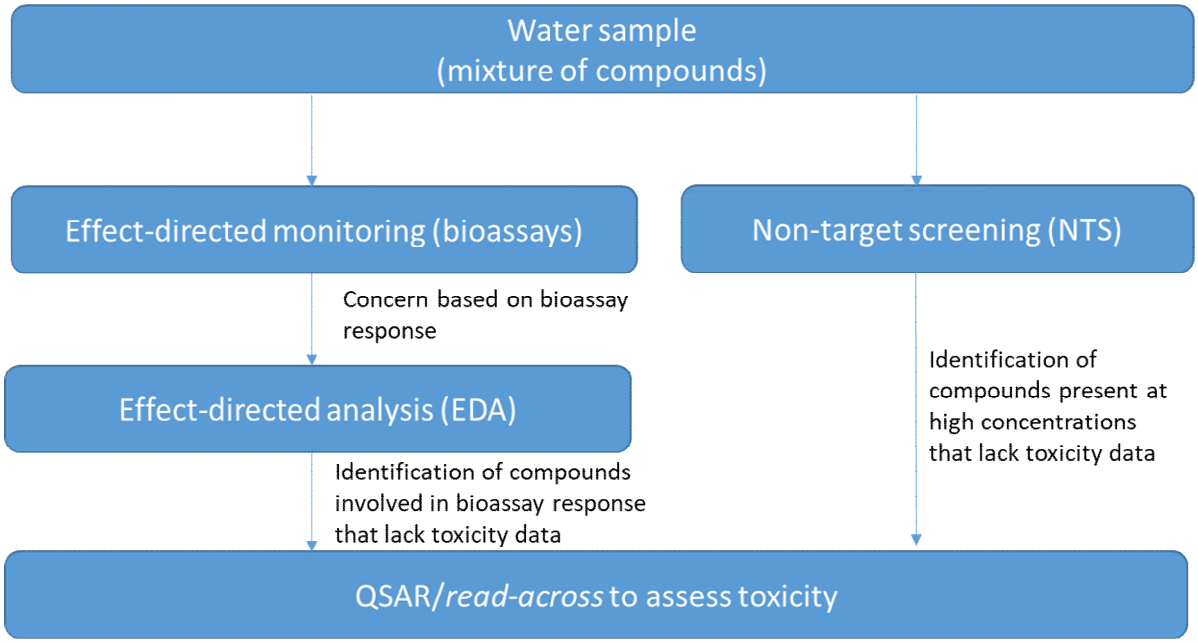


Figure 7: Schematic overview of the application of QSAR and read-across for prioritization of compounds for further research and interpretation of chemical analysis related to water quality monitoring.

When, based on effect-based measurements (bioassay responses), there is reason to further investigate possible risks to humans and the environment, effect-directed chemical analysis (EDA), in which water extracts are fractionated, can be used to specifically search for the chemical identity (substance name and chemical structure) of the substances that caused this response (Brekelmans et al. 2021; Houtman et al. 2020). For the identified substances for which toxicity data are missing, QSAR and read-across can be used to estimate the potential risks. Following non-target chemical screening (NTS) on a water sample, those substances that are present in high concentrations and for which toxicity data are lacking can be identified. QSAR and read-across can then be used to estimate potential risks to humans and the environment (Reus et al. 2022). Based on measured concentrations of chemicals in the water, the contribution to adverse effects to the human health or environment can be estimated for each chemical in the mixture. In this approach, the risk characterization ratio (RCR) is calculated by dividing the estimated daily intake of each chemical (based on measured concentrations) by the corresponding (provisional) drinking water guideline value ((p)GLV) (Rorije et al. 2022) or, alternatively, if a (p)GLV is not available, by the corresponding threshold of toxicological concern (TTC) (Kroes et al. 2005, Munro et al. 2008). Input for the TTC (information on whether a chemical is genotoxic and to which Cramer class it belongs) can be derived from the QSAR Toolbox and/or ToxTree and the relative contribution of individual chemicals can be related to the sum of the RCRs of all chemicals measured.

$$RCR = \frac{\text{Estimated exposure chemical 1} + \text{estimated exposure chemical 2} + \text{estimated exposure chemical X} + \ldots}{\text{GLV chemical 1} + \text{GLV chemical 2} + \text{TTC chemical X} + \ldots}$$

The variety of substances that emerge from EDA or NTS can thus be prioritized on this basis for quantitative concentration measurements (including chemical-analytical method development when necessary) and further risk assessment. Meekel et al (2021) developed a prioritization strategy in which fragment ion(s)/patterns of experimental fragmentation (MS2) spectra from NTS to structural alerts to identify the potential environmental or human health risk of substances in a mixture. For verification, results of the *in silico* tools (QSAR and read-across) can be subjected by testing the individual substances with bioassays (Reus et al. 2022).

Another important application of *in silico* models for water quality monitoring is to provide information for risk assessment of individual substances (Figure 8). When there is a quick need for an assessment of hazard and potential risks, such as when a data-poor substance is unexpectedly found in high concentrations in water, a read-across can be performed. Whereas the focus of prioritization for follow-up research needs is on the analysis and comparison of multiple substances simultaneously (the water sample containing a mixture), the focus of read-across is on individual substances. By means of an integrated approach in which various results from experimental studies and models are taken into account (weight of evidence), statements can be made about possible hazard and risks of a specific substance (Reus et al. 2022).

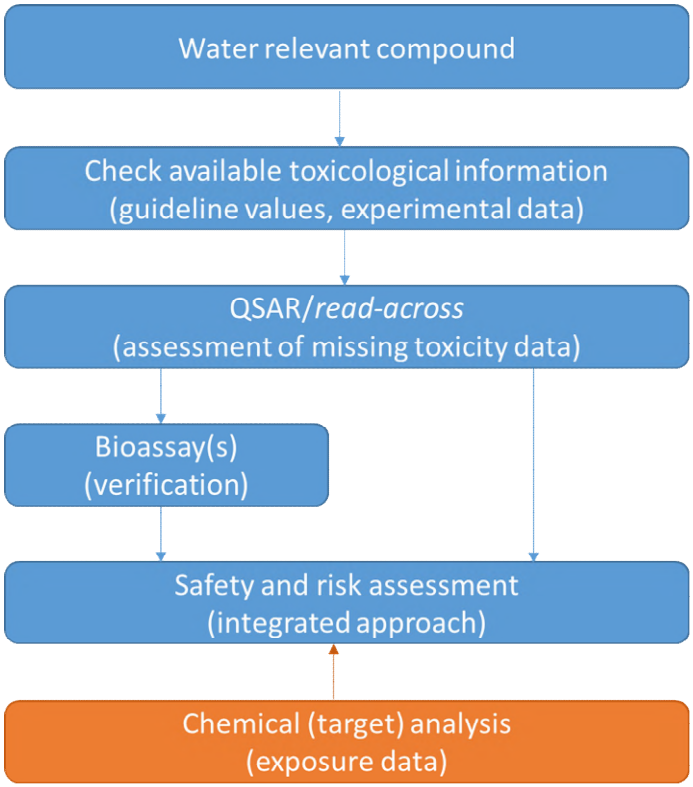


Figure 8: Schematic overview of the application of QSAR and read-across for safety and risk assessment of individual water relevant compounds.

7.2 Applicable *in silico* tools for water quality assessment

The QSAR Toolbox is suitable for collecting existing data of multiple endpoints, enabling rapid prioritization of individual or lists of chemicals based on existing data or structural alerts. The application is also suitable for hazard and risk assessment of individual chemicals that lack toxicological data using read across. It was concluded that the QSAR Toolbox has the advantage of manual grouping and offering many possibilities in terms of endpoints, but it has the disadvantage that it consequently induces variability among users and working with the software can be laborious. Compared to the QSAR Toolbox, working with the VEGA QSAR was found much more practical. The VEGA QSAR is suitable for rapid prioritization by simultaneous toxicity assessment of multiple endpoints for individual or lists of chemicals lacking toxicological data. CASE Ultra has more functionalities than the VEGA QSAR, but has the disadvantage of not being freely available. For all software applications, expert judgment based on toxicological and chemical background is needed to provide argumentation to accept or reject a toxicity prediction, especially when the data are being used for risk assessment (*e.g.* in case of emergencies).

Since most *in silico* models are user friendly, working with these models is not complicated. The interpretation of the results, however, requires expert judgement and basic knowledge on (organic) chemistry and toxicological mechanisms is essential. Justification is needed for acceptance or rejection of the prediction generated by the model and to verify if the model did select appropriate candidates for read across. For prioritization purposes, expert judgement may be limited, but it should be emphasized that the presence of structural alerts and the 'quick-and-dirty' hazard assessment does not necessarily mean that a substance indeed exerts this toxicological response in humans or in the environment. Moreover, even with expert judgement, a prediction does not necessarily mean that it is the truth. *In silico* toxicology can, however, conclude whether or not there is evidence for (absence of) a certain toxicological response or whether a certain response can (not) be excluded, based on the models applied.

Registration of all steps and choices is essential for transparency and for verification of the results (QC/QA). In line with the ICH M7 guideline (EMA 2015), for robust predictions it is recommended to include at least one expert rule-based and one statistical-based model for the toxicity prediction of water relevant substances. Although the QSAR Toolbox was considered the most robust *in silico* model studied, it only provides an expert-rule based model and using the QSAR Toolbox only is therefore not sufficient for hazard and risk assessment. VEGA QSAR and CASE Ultra provide statistical-based models that can support and confirm the prediction from QSAR Toolbox. In case of emergencies, it is recommended to use the QSAR Toolbox for a robust toxicity prediction of the chemical(s) of interest. The VEGA QSAR can be used as a quick-scan and second model. To increase the robustness of the prediction, it is however recommended to use CASE Ultra as second model as it includes more complete databases than VEGA QSAR.

The current research focused on the endpoints mutagenicity and estrogenic activity. Although the QSAR Toolbox has proven its applicability for prediction of bacterial mutagenicity, the applicability for prediction of estrogenic activity was considered limited due to lack of analogues with data on ToxCast ERα reporter gene assays for several of the water-relevant chemicals studied. VEGA QSAR was able to predict estrogenic activity in a qualitative manner, but not all analogues suggested by VEGA QSAR were considered equally relevant. Taken these limitations and uncertainties into account, the relevance of prediction of estrogenic activity using the current *in silico* tools used is questionable. Further development of the tools, including more chemicals with data is needed for reliable predictions on estrogenic activity.

No other endpoints than bacterial mutagenicity and estrogenic activity were studied because the current project focused on the combination of using *in silico* tools and bioassays, and these were the *in vitro* bioassays available for water quality assessment that showed the most prominent link to toxicological endpoints in the software. The *in silico* tools are mainly focused on endpoints required for hazard and risk assessment in a regulatory context. Implementation of the *in vitro* micronucleus test for water quality assessment (ISO 21427-2:2006, Reifferscheid et al. 2008), however, can be considered as a second bioassay for assessment of DNA damage other than mutagenicity. The micronucleus test detects chromosomal aberrations such as loss of complete chromosomes or parts thereof, which result in a small nucleus (micronucleus) in addition to the main nucleus. This type of DNA damage is different to mutagenicity as detected with the Ames fluctuation test. Both the QSAR Toolbox and VEGA QSAR allow prediction of genotoxicity using the micronucleus test. The Umu test (ISO 13829:2000, Oda et al. 1985) can be used to predict genotoxicity as part of water quality assessment, but neither in the QSAR Toolbox nor in the VEGA QSAR there is an endpoint that exactly matches with the Umu test. The p53 CALUX (van der Linden et al. 2014) could be linked to ToxCast bioassays based on the assay principle, however the availability of data for analogue chemicals may yet be limited. Alternatively, p53 CALUX and Umu test results can be linked to the genotoxicity tests available in the QSAR Toolbox and VEGA QSAR (*i.e.* bacterial mutagenicity, mammalian cell mutagenicity, micronucleus test, comet assay and chromosomal abberation test), but the concordance of the results and thus reliability of the prediction warrants further investigation.

Other water relevant toxicological endpoints, such as neurotoxicity and reproductive toxicity, are complex processes in which many different pathways are involved. If all those pathways would be incorporated in an *in silico* tool this might be an interesting application, since investigation of all those pathways with bioassays is difficult, time-consuming and labour-intensive. Moreover, at the moment there are no standardized bioassays (*in vitro* or using non-vertebrates) available for neurotoxicity (BTO 2020.035). In addition, the lack of available experimental data on regulatory animal studies, hampers reliable applicability of *in silico* tools for the prediction of neurotoxicity and reproductive toxicity. The same holds true for carcinogenicity. Further development of the *in silico* tools, including more chemicals with data is needed for reliable predictions on these endpoints.

7.3 Discussion of experimental results

In the current research, the predicted responses on bacterial mutagenicity showed good concordance with the Ames fluctuation experimental results. It should be noted, however, that this conclusion is based on a limited number of compounds and that only negative responses were obtained. At this point it cannot be excluded that the prediction of positive bioassay responses induced by mutagenic substances is equally reliable. It is thus needed to investigate additional compounds, including chemicals for which positive responses are expected. The current chemical list did not include such chemicals.

For some chemicals, it was not possible to reliably predict the mutagenic response especially with VEGA QSAR. In these cases, the prediction was regarded as inconclusive for the specific model, most often due to lack of relevant analogues for read-across. Since it is recommended to use at least two different models (*i.e.,* expert rule-based and statistical-based), it is possible that an inconclusive prediction in one model results in an overall inconclusive prediction for the chemical. It is also possible that the results of different models are contradictory. In case of inconclusive or contradictory results, experimental studies on the endpoint of interest are needed to draw conclusions on hazard assessment.

Since some compounds require metabolic activation to exert their genotoxic effects (Guengerich, 2008; Nebert et al. 2006), the formation of metabolites can be predicted in the QSAR Toolbox using the following profilers relevant to mammalian metabolism (hydrolysis and microbial metabolism were not take into account): 1) Observed mammalian metabolism, 2) Observed Rat *In vivo* metabolism, 3) Observed rat liver metabolism with quantitative data, 4) Observed Rat Liver S9 metabolism, 5) *In vivo* Rat metabolism simulator, 6) Rat liver S9 metabolism simulator.
Investigation of metabolites is in principle only required for compounds that were subjected to the read-across (*i.e.,* the compounds for which experimental data was lacking). For a comprehensive risk assessment, all (predicted) metabolites of the target compound should be assessed individually for the endpoint of interest (not part of the present study).

The results of the Ames fluctuation experiments show that it is not easy to select appropriate dose levels. In this research, maximum concentrations of the first experiment were selected based on either limit concentrations generally used for *in vitro* assays or limited by solubility, but the selected concentrations appeared to be cytotoxic. The advantage of observing cytotoxicity is that it can be assumed that the concentration levels were sufficiently high to induce a genotoxic response, but the disadvantage is that excessive cytotoxicity may lead to false negative responses and that consequently no conclusion on mutagenicity can be drawn, resulting in the need for a repeat experiment. Performance of a dose-range finding experiment and/or increasing the number of concentration levels in the initial experiment reduce the risk of the need for repeating an experiment, yet it is less time-efficient than the pragmatic approach used in this study.

7.4 Dempster Shafer Theory (DST)

In the previous chapters we showed that we used multiple sources of evidence (*in silico* tools) to predict mutagenicity for compounds lacking toxicological data. Each of these computational tools relies on approaches such as QSAR and read-across. However, most of these tools lack information on the uncertainty of the model outcomes resulting from limitations in the model itself. Although each model may work well to predict mutagenicity of compounds, each model has its own limitations and therefore may lack reliability on its own. Weight-of-evidence approaches combining predictions from various methods (not solely including *in silico* tools) have become commonly accepted (ECHA, 2023; EFSA, 2017).

One of such weight-of-evidence approaches is the Dempster Shafer Theory (DST). A technical explanation on the DST theory is provided in Appendix II.

Future research on how toxicological predictions with varying uncertainty can be combined by using DST is ongoing (Rathman et al., 2018). The main challenge is to reliably estimate prediction uncertainty based on results from multiple models, information on the uncertainty of each model prediction is required. Although both VEGA QSAR and CASE Ultra provide information on the reliability of each model predictions, only CASE Ultra provides statistically derived probability values, which makes this model particularly useful when applying DST. To account for uncertainties and variability in the different model outcomes, more research into application and acceptance of DST in (*in silico*) toxicological assessment is required (Benigni et al., 2019).

Conclusion

In silico tools provide insight into which micropollutants in water pose a potential risk to health and the environment. Therefore, they have a clear added value as a solution for different water quality questions around prioritization for research needs and safety and risk assessment. Practical measures and decision-making to mitigate any risks can thus be purposefully chosen and deployed. Further development of in silico tools with regards to endpoints other than genotoxicity is needed.

8 Knowledge gaps and recommendations for future research

The current project focused on the application of *in silico* models in combination with bioassays on individual water relevant chemicals. A next step would be to integrate *in silico* toxicology in a strategy for hazard assessment of water samples with advanced chemical analysis and identification using high resolution mass spectrometry to demonstrate the applicability of *in silico* toxicology in practice. Ideally, this is done in a multidisciplinary collaboration such as proposed in a pending research proposal called Waterpatroon (*Efficient risk identification of water quality by integrating chemical and toxicological data,* risk-based prioritization based on MS2 spectra) and/or in connection to other (BTO) projects in which non-target screening and/or bioassays are performed.

For ongoing projects and hazard and risk assessment activities, the knowledge obtained in the current project on the use of *in silico* models is already being applied. Recently, substances for studies on treatment efficiency were prioritized based on the presence of structural alerts for water-relevant toxicological endpoints derived from profiling in the QSAR Toolbox. In addition, *in silico* tools are being used to support toxicological risk assessments for contaminants found in drinking water sources and to evaluate toxicological risk assessments based on *in silico* predictions (peer-review). However, more in depth knowledge on the use and acceptance of *in silico* models in a regulatory context is needed for a more accurate hazard and risk assessment in a water-relevant context. This can be achieved by systematically reviewing and summarizing existing guidance documents from the European Food Safety Authority (EFSA), ECHA and OECD, and by discussing this with external experts.

Xenobiotic metabolism was considered only to a limited extent in the current project because the selected compounds overall showed negative responses, both in the presence and absence of metabolic enzymes (S9). Nevertheless, it is known that metabolism is an important aspect in genotoxicity assessment as some compounds are progenotoxins, *i.e.,* these compounds become genotoxic after metabolic activation. In follow-up research, consideration of metabolism can be integrated in the *in silico* strategy, and the opinion of authorities should also be included.

The Dempster Shafer Theory (paragraph 7.4) was explored in the current project, but further elaboration on the application in hazard assessment using *in silico* toxicology and experimental data is warranted. It is recommended that the applicability of DST or other weight-of-evidence approaches is further evaluated.

Automatization, *e.g.,* using workflows or programming, was not part of the current project. However, automatization was explored and it is foreseen that this can be expanded in follow-up research. For VEGA QSAR, this can be done using KNIME[22]. For the QSAR Toolbox, programming features are not embedded in the software application itself. However, output of the QSAR Toolbox can be generated in Excel and Excel files can be coupled to R scripts for further data processing.

The Ames fluctuation test result of Levonorgestrel was equivocal. With additional experiments the conclusion may become more definitive.

[22] Konstanz Information Miner (KNIME) is a free and open-source data analytics, reporting and integration platform. KNIME integrates various components for machine learning and data mining through its modular data pipelining "Building Blocks of Analytics" concept. A graphical user interface and use of JDBC allows assembly of nodes blending different data sources, including preprocessing (ETL: Extraction, Transformation, Loading), for modeling, data analysis and visualization without, or with only minimal, programming (en.wikipedia.org).

9 Dissemination

The developed strategies on application of *in silico* tools for water quality assessment have been published in H2O in March 2022 (Reus et al. 2022). A scientific publication on the combined application of *in silico* tools and bioassays for water quality assessment is foreseen after investigating water samples in addition to individual substances. It is foreseen that chemical analysis (non-targeted) is also included in this follow-up research.

Combined results of the Ames fluctuation test and the read-across, as well as the proposed strategies for water quality assessment are presented and discussed at the International Congress on Toxicology (ICT 2022), held in Maastricht, September 18-21. This included both a short oral communication and an e-poster.

References

Baan RA, Stewart BW, Straif K, editors. Tumour Site Concordance and Mechanisms of Carcinogenesis. Lyon (FR): International Agency for Research on Cancer; 2019.

Baken KA, Sjerps RMA, Schriks M, van Wezel AP. Toxicological risk assessment and prioritization of drinking water relevant contaminants of emerging concern. Environ Int. 2018 Sep;118:293-303.

de Barros ALC, da Silva Rodrigues DA, da Cunha CCRF, Chagas IASD, Santo DRDE, Silva SQ, Afonso RJCF. Aqueous chlorination of herbicide metribuzin: Identification and elucidation of "new" disinfection by-products, degradation pathway and toxicity evaluation. Water Res. 2021 Feb 1;189:116545.

Basu AK. DNA Damage, Mutagenesis and Cancer. Int J Mol Sci. 2018 Mar 23;19(4):970.

Been F, Kruve A, Vughs D, Meekel N, Reus A, Zwartsen A, Wessel A, Fischer A, Ter Laak T, Brunner AM. Risk-based prioritization of suspects detected in riverine water using complementary chromatographic techniques. Water Res. 2021 Oct 1;204:117612.

Benfenati E, Manganaro A, Gini G. VEGA-QSAR: AI inside a platform for predictive toxicology, Proceedings of the workshop “Popularize Artificial Intelligence 2013”. December 5th 2013, Turin, Italy. Published on CEUR Workshop Proceedings Vol-1107, paper_8 (ceur-ws.org)

Benfenati E, Chaudhry Q, Gini G, Dorne JL. Integrating in silico models and read-across methods for predicting toxicity of chemicals: A step-wise strategy. Environ Int. 2019 Oct;131:105060.

Benigni R. Chemical structure of mutagens and carcinogens and the relationship with biological activity. J Exp Clin Cancer Res. 2004 Mar;23(1):5-8.

Benigni R, Bossa C. Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. Mutat Res. 2008 Sep-Oct;659(3):248-61.

Benigni R, Serafimova R, Parra Morte JM, Battistelli CL, Bossa C, Giuliani A, Fioravanzo E, Bassan A, Gatnik MF, Rathman J, Yang C, Mostrag-Szlichtyng A, Sacher O, Tcheremenskaia O. Evaluation of the applicability of existing (Q)SAR models for predicting the genotoxicity of pesticides and similarity analysis related with genotoxicity of pesticides for facilitating of grouping and read across: An EFSA funded project. Regul Toxicol Pharmacol. 2020 Jul;114:104658.

Benigni R, Bossa C, Tcheremenskaia O. Nongenotoxic carcinogenicity of chemicals: mechanisms of action and early recognition through a new set of structural alerts. Chem Rev. 2013 May 8;113(5):2940-57..

Brack W, Ait-Aissa S, Burgess RM, Busch W, Creusot N, Di Paolo C, Escher BI, Mark Hewitt L, Hilscherova K, Hollender J, Hollert H, Jonker W, Kool J, Lamoree M, Muschket M, Neumann S, Rostkowski P, Ruttkies C, Schollee J, Schymanski EL, Schulze T, Seiler TB, Tindall AJ, De Aragão Umbuzeiro G, Vrana B, Krauss M. Effect-directed analysis supporting monitoring of aquatic environments--An in-depth overview. Sci Total Environ. 2016 Feb 15;544:1073-118.

Brekelmans S, Slootweg T, Houtman C. Effect Directed Analysis achterhaalt verantwoordelijke stoffen voor bioassayrespons. H2O-online, 8 oktober 2021

Brunner AM, Vughs D, Siegers W, Bertelkamp C, Hofman-Caris R, Kolkman A, Ter Laak T. Monitoring transformation product formation in the drinking water treatments rapid sand filtration and ozonation. Chemosphere. 2019 Jan;214:801-811.

Brunner AM, Bertelkamp C, Dingemans MML, Kolkman A, Wols B, Harmsen D, Siegers W, Martijn BJ, Oorthuizen WA, Ter Laak TL. Integration of target analyses, non-target screening and effect-based monitoring to assess OMP related water quality changes in drinking water treatment. Sci Total Environ. 2020 Feb 25;705:135779.

BTO 2016.069, The Threshold of Toxicological Concern (TTC): refinement of the concept and application to drinking water.

BTO 2018.030, Tools for human health risk assessment of emerging chemicals. www.kwrwater.nl

BTO 2020.035, Bioassays voor neuroactieve stoffen in water. www.kwrwater.nl

Carpinteiro I, Rodil R, Quintana JB, Cela R. Reaction of diazepam and related benzodiazepines with chlorine. Kinetics, transformation products and in-silico toxicological assessment. Water Res. 2017 Sep 1;120:280-289.

Chakravarti SK, Saiakhov RD, Klopman G. Optimizing predictive performance of CASE Ultra expert system models using the applicability domains of individual toxicity alerts. J Chem Inf Model. 2012 Oct 22;52(10):2609-18.

Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. Environ Mol Mutagen. 2017 Jun;58(5):235-263.

Chen M, Wei D, Wang F, Yin J, Li M, Du Y. Bioassay- and QSAR-based screening of toxic transformation products and their formation under chlorination treatment on levofloxacin. J Hazard Mater. 2021 Jul 15;414:125495.Cronin MTD, Enoch SJ, Mellor CL, Przybylak KR, Richarz AN, Madden JC. *In Silico* Prediction of Organ Level Toxicity: Linking Chemistry to Adverse Effects. Toxicol Res. 2017 Jul;33(3):173-182

Cvetnic M, Juretic Perisic D, Kovacic M, Ukic S, Bolanca T, Rasulev B, Kusic H, Loncaric Bozic A. Toxicity of aromatic pollutants and photooxidative intermediates in water: A QSAR study. Ecotoxicol Environ Saf. 2019 Mar;169:918-927.

Danishuddin, Khan AU. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. Drug Discov Today. 2016 Aug;21(8):1291-302.

Dingemans MM, Baken KA, van der Oost R, Schriks M, van Wezel AP. Risk-based approach in the revised European Union drinking water legislation: Opportunities for bioanalytical tools. Integr Environ Assess Manag. 2019 Jan;15(1):126-134.

Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ. The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci. 2007 Jan;95(1):5-12.

ECHA. Guidance on information requirements and chemical safety assessment Chapter R.6: QSARs and grouping of chemicals. May 2008, RDDS - Batch Process 0.9 (europa.eu)

ECHA. Practical guide: How to use and report (Q)SARs. Version 3.1 – July 2016. https://echa.europe.eu

ECHA. Weight of evidence. Weight of evidence - ECHA (europa.eu). Accessed on January 2, 2023.

EFSA. Scientific Opinion on genotoxicity testing strategies applicable to food and feed safety assessment. EFSA Journal 2011;9(9), 2379.

EFSA. Guidance on the use of the weight of evidence approach in scientific assessments. EFSA Journal 2017;15(8):4971

Ellison CM, Madden JC, Judson P, Cronin MT. Using In Silico Tools in a Weight of Evidence Approach to Aid Toxicological Assessment. Mol Inform. 2010 Jan 12;29(1-2):97-110.

EMA. ICH guideline M7(R1) on assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk. EMA/CHMP/ICH/83812/2013, 25 August 2015.

Friedberg EC, McDaniel LD, Schultz RA. The role of endogenous and exogenous DNA damage and mutagenesis. Curr Opin Genet Dev. 2004 Feb;14(1):5-10.

Guengerich FP. Cytochrome p450 and chemical toxicology. Chem Res Toxicol. 2008 Jan;21(1):70-83.

Han Y, Ma M, Li N, Hou R, Huang C, Oda Y, Wang Z. Chlorination, chloramination and ozonation of carbamazepine enhance cytotoxicity and genotoxicity: Multi-endpoint evaluation and identification of its genotoxic transformation products. J Hazard Mater. 2018 Jan 15;342:679-688.

Hewitt M, Ellison CM, Enoch SJ, Madden JC, Cronin MT. Integrating (Q)SAR models, expert systems and read-across approaches for the prediction of developmental toxicity. Reprod Toxicol. 2010 Aug;30(1):147-60.

Houtman CJ, ten Broek R, de Jong K, Pieterse B, Kroesbergen J. A multicomponent snapshot of pharmaceuticals and pesticides in the river Meuse basin. Environ Toxicol Chem. 2013 Nov;32(11):2449-59.

Houtman CJ, Ten Broek R, van Oorschot Y, Kloes D, van der Oost R, Rosielle M, Lamoree MH. High resolution effect-directed analysis of steroid hormone (ant)agonists in surface and wastewater quality monitoring. Environ Toxicol Pharmacol. 2020 Nov;80:103460.

ISO 11350:2012; Water quality — Determination of the genotoxicity of water and waste water — Salmonella/microsome fluctuation test (Ames fluctuation test), June 2012.

ISO 13829:2000; Water quality — Determination of the genotoxicity of water and waste water using the umu-test, March 2000.

ISO 21427-2:2006; Water quality — Evaluation of genotoxicity by measurement of the induction of micronuclei — Part 2: Mixed population method using the cell line V79, November 2016.

Kegel FS, Rietman BM, Verliefde AR. Reverse osmosis followed by activated carbon filtration for efficient removal of organic micropollutants from river bank filtrate. Water Sci Technol. 2010;61(10):2603-10.

Kroes R, Kleiner J, Renwick A. The threshold of toxicological concern concept in risk assessment. Toxicol Sci. 2005 Aug;86(2):226-30.

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

Lu BQ, Liu SS, Wang ZJ, Xu YQ. Conlecs: A novel procedure for deriving the concentration limits of chemicals outside the criteria of human drinking water using existing criteria and species sensitivity distribution based on quantitative structure-activity relationship prediction. J Hazard Mater. 2020 Feb 15;384:121380.

Mahmoud WM, Toolaram AP, Menz J, Leder C, Schneider M, Kümmerer K. Identification of phototransformation products of thalidomide and mixture toxicity assessment: an experimental and quantitative structural activity relationships (QSAR) approach. Water Res. 2014 Feb 1;49:11-22.

Martin MT, Judson RS, Reif DM, Kavlock RJ, Dix DJ. Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef Database. Environ Health Perspect. 2009 Mar;117(3):392-9.

Munro IC, Renwick AG, Danielewska-Nikiel B. The Threshold of Toxicological Concern (TTC) in risk assessment. Toxicol Lett. 2008 Aug 15;180(2):151-6.

Matsushita T, Honda S, Kuriyama T, Fujita Y, Kondo T, Matsui Y, Shirasaki N, Takanashi H, Kameya T. Identification of mutagenic transformation products generated during oxidation of 3-methyl-4-nitrophenol solutions by orbitrap tandem mass spectrometry and quantitative structure-activity relationship analyses. Water Res. 2018 Feb 1;129:347-356.

Matsushita T, Hashizuka M, Kuriyama T, Matsui Y, Shirasaki N. Use of orbitrap-MS/MS and QSAR analyses to estimate mutagenic transformation products of iopamidol generated during ozonation and chlorination. Chemosphere. 2016 Apr;148:233-40.

Meekel N, Vughs D, Béen F, Brunner AM. Online Prioritization of Toxic Compounds in Water Samples through Intelligent HRMS Data Acquisition. Anal Chem. 2021 Mar 30;93(12):5071-5080.

Mestankova H, Schirmer K, Canonica S, von Gunten U. Development of mutagenicity during degradation of N-nitrosamines by advanced oxidation processes. Water Res. 2014 Dec 1;66:399-410.

Moon J, Lee B, Ra JS, Kim KT. Predicting PBT and CMR properties of substances of very high concern (SVHCs) using QSAR models, and application for K-REACH. Toxicol Rep. 2020 Aug 15;7:995-1000.

Nebert DW, Dalton TP. The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. Nat Rev Cancer. 2006 Dec;6(12):947-60.

Nohmi T. Thresholds of Genotoxic and Non-Genotoxic Carcinogens. Toxicol Res. 2018 Oct;34(4):281-290.

Oda Y, Nakamura S, Oki I, Kato T, Shinagawa H. Evaluation of the new system (umu-test) for the detection of environmental mutagens and carcinogens. Mutat Res. 1985 Oct;147(5):219-29.

Pérez-Garrido A, González MP, Escudero AG. Halogenated derivatives QSAR model using spectral moments to predict haloacetic acids (HAA) mutagenicity. Bioorg Med Chem. 2008 May 15;16(10):5720-32.

Phillips DH, Arlt VM. Genotoxicity: damage to DNA and its consequences. EXS. 2009;99:87-110.

Plošnik A, Vračko M, Dolenc MS. Mutagenic and carcinogenic structural alerts and their mechanisms of action. Arh Hig Rada Toksikol. 2016 Sep 1;67(3):169-182.

Qin L, Zhang X, Chen Y, Mo L, Zeng H, Liang Y. Predictive QSAR Models for the Toxicity of Disinfection Byproducts. Molecules. 2017 Oct 9;22(10):1671.

Rathman JF, Yang C, Zhou, H. Dempster-Shafer theory for combining *in silico* evidence and estimating uncertainty in chemical risk assessment. Comput Toxicol. 2018, May:6,:16-31.

Reifferscheid G, Maes HM, Allner B, Badurova J, Belkin S, Bluhm K, Brauer F, Bressling J, Domeneghetti S, Elad T, Flückiger-Isler S, Grummt HJ, Gürtler R, Hecht A, Heringa MB, Hollert H, Huber S, Kramer M, Magdeburg A, Ratte HT, Sauerborn-Klobucar R, Sokolowski A, Soldan P, Smital T, Stalter D, Venier P, Ziemann C, Zipperle J, Buchinger S. International round-robin study on the Ames fluctuation test. Environ Mol Mutagen. 2012 Apr;53(3):185-97.

Reus A, Dingemans M, Hoondert R, Houtman C, van der Oost R. QSAR en read-across-modellen voor waterkwaliteit. H2O-online, 18 maart 2022.

Roberts G, Myatt GJ, Johnson WP, Cross KP, Blower PE Jr. LeadScope: software for exploring large sets of screening data. J Chem Inf Comput Sci. 2000 Nov-Dec;40(6):1302-14.

Robitaille J, Denslow ND, Escher BI, Kurita-Oyamada HG, Marlatt V, Martyniuk CJ, Navarro-Martín L, Prosser R, Sanderson T, Yargeau V, Langlois VS. Towards regulation of Endocrine Disrupting chemicals (EDCs) in water resources using bioassays - A guide to developing a testing strategy. Environ Res. 2022 Apr 1;205:112483. doi: 10.1016/j.envres.2021.112483

Rorije E, Wassenaar PNH, Slootweg J, van Leeuwen L, van Broekhuizen FA, Posthuma L. Characterization of ecotoxicological risks from unintentional mixture exposures calculated from European freshwater monitoring data: Forwarding prospective chemical risk management. Sci Total Environ. 2022 May 20;822:153385.

Rozemeijer JC, Broers HP. The groundwater contribution to surface water contamination in a region with intensive agricultural land use (Noord-Brabant, The Netherlands). Environ Pollut. 2007 Aug;148(3):695-706. doi: 10.1016/j.envpol.2007.01.028.

Sanabria P, Scunderlick D, Wilde ML, Lüdtke DS, Sirtori C. Solar photo-Fenton treatment of the anti-cancer drug anastrozole in different aqueous matrices at near-neutral pH: Transformation products identification, pathways proposal, and in silico (Q)SAR risk assessment. Sci Total Environ. 2021 Feb 1;754:142300.

Schipper PN, Vissers MJ, van der Linden AM. Pesticides in groundwater and drinking water wells: overview of the situation in the Netherlands. Water Sci Technol. 2008;57(8):1277-86.

Schriks M, Baken K, Simon E, Besselink H, Van der Linden S, Kienle C, Van der Burg B. Selection criteria to select in vitro bioassays for implementation and use. DEMEAU (FP7) report, 2015. URL: http://demeau-fp7.eu/content/d411-selection-criteria-select-vitro-bioassays-implementation-and-use

Shah I, Liu J, Judson RS, Thomas RS, Patlewicz G. Systematically evaluating read-across prediction and performance using a local validity approach characterized by chemical structure and bioactivity information. Regul Toxicol Pharmacol. 2016 Aug;79:12-24.

Shao Y, Hollert H, Tarcai Z, Deutschmann B, Seiler TB. Integrating bioassays, chemical analysis and in silico techniques to identify genotoxicants in surface water. Sci Total Environ. 2019 Feb 10;650(Pt 2):3084-3092.

Sharma A, Ahmad J, Flora SJS. Application of advanced oxidation processes and toxicity assessment of transformation products. Environ Res. 2018 Nov;167:223-233.

Simon-Hettich B, Rothfuss A, Steger-Hartmann T. Use of computer-assisted prediction of toxic effects of chemical substances. Toxicology. 2006 Jul 5;224(1-2):156-62.

Ter Laak TL, Puijker LM, van Leerdam JA, Raat KJ, Kolkman A, de Voogt P, van Wezel AP. Broad target chemical screening approach used as tool for rapid assessment of groundwater quality. Sci Total Environ. 2012 Jun 15;427-428:308-13.

van der Linden SC, von Bergh AR, van Vught-Lussenburg BM, Jonker LR, Teunis M, Krul CA, van der Burg B. Development of a panel of high-throughput reporter-gene assays to detect genotoxicity and oxidative stress. Mutat Res Genet Toxicol Environ Mutagen. 2014 Jan 15;760:23-32

Verliefde A, Cornelissen E, Amy G, Van der Bruggen B, van Dijk H. Priority organic micropollutants in water sources in Flanders and the Netherlands and assessment of removal possibilities with nanofiltration. Environ Pollut. 2007 Mar;146(1):281-9.

Wei X, Yang M, Zhu Q, Wagner ED, Plewa MJ. Comparative Quantitative Toxicology and QSAR Modeling of the Haloacetonitriles: Forcing Agents of Water Disinfection Byproduct Toxicity. Environ Sci Technol. 2020 Jul 21;54(14):8909-8918.

Wichard JD. In silico prediction of genotoxicity. Food Chem Toxicol. 2017 Aug;106(Pt B):595-599.

Yang RS, Thomas RS, Gustafson DL, Campain J, Benjamin SA, Verhaar HJ, Mumtaz MM. Approaches to developing alternative and predictive toxicology based on PBPK/PD and QSAR modeling. Environ Health Perspect. 1998 Dec;106 Suppl 6(Suppl 6):1385-93.

Ye Y, Weiwei J, Na L, Mei M, Donghong W, Zijian W, Kaifeng R. Assessing of genotoxicity of 16 centralized source-waters in China by means of the SOS/umu assay and the micronucleus test: initial identification of the potential genotoxicants by use of a GC/MS method and the QSAR Toolbox 3.0. Mutat Res Genet Toxicol Environ Mutagen. 2014 Mar 15;763:36-43.

Yordanova D, Schultz TW, Kuseva C, Tankova K, Ivanova H, Dermen I, Pavlov T, Temelkov S, Chapkanov A, Georgiev M. Automated and standardized workflows in the OECD QSAR Toolbox. Comput Toxicol. 2019 May;10:89-104.

Zhang Z, Zhu Q, Huang C, Yang M, Li J, Chen Y, Yang B, Zhao X. Comparative cytotoxicity of halogenated aromatic DBPs and implications of the corresponding developed QSAR model to toxicity mechanisms of those DBPs: Binding interactions between aromatic DBPs and catalase play an important role. Water Res. 2020 Mar 1;170:115283.

I Water relevant compounds of interest for toxicological evaluation

Name	CAS	SMILES	Origin	Reference
1-(2-benzaldehyde)-(1H, 3H)-quinazoline-2,4-dione	unknown	unknown	transformation product	Brunner et al. 2019
1-(2-benzoic acid)-(1H, 3H)-quinazoline-2,4-dione	unknown	unknown	transformation product	Brunner et al. 2019
1-(2-benzoic acid)-4-hydro(1H, 3H)-quinazoline-2-one	unknown	unknown	transformation product	Brunner et al. 2019
1,3-Propanedione, 1,3-diphenyl-	120-46-7	O=C(CC(=O)C1=CC=CC=C1)C1=CC=CC=C1	industrial	Baken et al. 2018
1H-Benzotriazole, 5,6-dimethyl-	4184-79-6	CC1=CC2=NNN=C2C=C1C	industrial	Baken et al. 2018
1H-Benzotriazole, 5-methyl-	136-85-6	CC1=CC2=C(NN=N2)C=C1	industrial	Baken et al. 2018
1H-Indene, 2,3-dihydro-	496-11-7	C1CC2=CC=CC=C2C1	industrial	Baken et al. 2018
2-(Diethylamino)ethanol	100-37-8	CCN(CC)CCO	transformation product	Unpublished data KWR
2-Aminobenzoic acid	118-92-3 / 1321-11-5 / 93762-40-4	Nc1ccccc1C(O)=O	transformation product	Unpublished data KWR
2-Isopropoxy-1,3-cyclohexadiene	98677-91-9	CC(C)OC1=CC-CC=C1	transformation product	Unpublished data KWR
2-Propenoic acid/2-acrylamido-2-methyl-1-propanesulfonic acid copolymer natriumzout met CAS nr 77019-71-7	40623-75-4	OC(=O)C=C.CCC(C)(N=C(O)C=C)S(O)(=O)=O	industrial	Present in drinking water sources
3-Pyrazolpropanoic acid	89532-73-0	OC(=O)CCN1C=CC=N1	industrial	Present in drinking water sources

Name	CAS	SMILES	Origin	Reference
3-Pyrazolpropionitrile	88393-88-8	N#CCCN1C=CC=N1	industrial	Present in drinking water sources
4-OH-omeprazole sulfide	103876-98-8	COC1=CC2=C(C=C1)N=C(N2)SCC1=C(C)C(=O)C(C)=CN1	pharmaceutical metabolite	Present in drinking water sources
9-Hydroxy-acridine			transformation product	Brunner et al. 2019
Acridone	578-95-0	O=C1C2=CC=CC=C2NC2=CC=CC=C12	transformation product	Brunner et al. 2019
Adenosine 5 -(tetrahydrogen triphosphate) [ATP]	56-65-5	NC1=NC=NC2=C1N=CN2[C@@H]1O[C@H](COP(O)(=O)OP(O)(=O)OP(O)(O)=O)[C@@H](O)[C@H]1O	natural origin	Baken et al. 2018
Alpha-hydroxyisobutyric acid	594-61-6	CC(C)(O)C(O)=O	transformation product	Brunner et al. 2019
Amantadine	768-94-5	NC12CC3CC(CC(C3)C1)C2	pharmaceutical	Present in drinking water sources
Anabasine	40774-73-0	-	natural origin	Present in drinking water sources
Antraquinone	84-65-1	O=C1C2=C(C=CC=C2)C(=O)C2=C1C=CC=C2	industrial	Present in drinking water sources
Benzamide, 2-amino-N-(1-methylethyl)-	30391-89-0	CC(C)NC(=O)C1=C(N)C=CC=C1	pesticide metabolite	Baken et al. 2018
Benzyl butyrate	103-37-7	CCCC(=O)OCC1=CC=CC=C1	transformation product	Unpublished data KWR
Bis(4-ethylbenzylidene)sorbitol	79072-96-1	CCC1=CC=C(C=C1)C1OCC2OC(OC(C(O)CO)C2O1)C1=CC=C(CC)C=C1	transformation product	Unpublished data KWR
Clotrimazole	23593-75-1	ClC1=C(C=CC=C1)C(N1C=CN=C1)(C1=CC=CC=C1)C1=CC=CC=C1	pharmaceutical	Present in drinking water sources
Cotinine	486-56-6	CN1[C@@H](CCC1=O)C1=CN=CC=C1	nicotine metabolite	Present in drinking water sources
Cyaanureum	2208-89-1	OC(=N)NC#N	industrial	Present in drinking water sources
Cyanopropanal	3515-93-3	O=CCCC#N	industrial	Present in drinking water sources
Cyanopropanalcyanohydrin	2478-49-1	OC(CCC#N)C#N	industrial	Present in drinking water sources

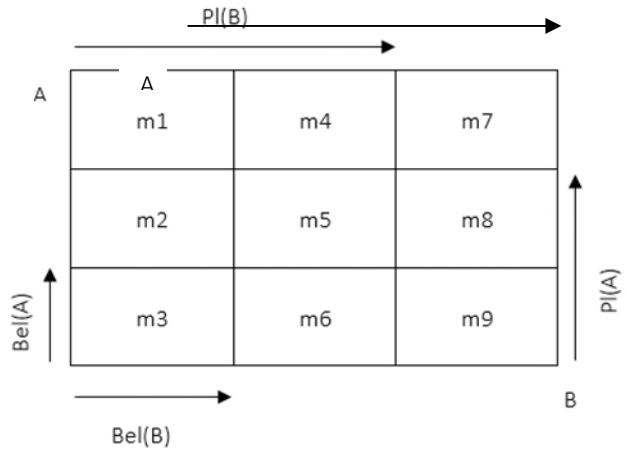
Name	CAS	SMILES	Origin	Reference
Cyclohexane	110-82-7	C1CCCCC1	industrial	Baken et al. 2018
Deschlormetolachlor	126605-22-9	CCC1=CC=CC(C)=C1N(C(C)COC)C(C)=O	transformation product	Brunner et al. 2019
Desmetramadol	73986-53-5	CN(C)CC1CCCCC1(O)C1=CC(O)=CC=C1	pharmaceutical metabolite	Present in drinking water sources
Diglyme	111-96-6	COCCOCCOC	industrial	Present in drinking water sources
Dinoterb	1420-07-1	CC(C)(C)C1=CC(=CC(=C1O)[N+]([O-])=O)[N+]([O-])=O	pesticide	Baken et al. 2018
EDTA	60-00-4	OC(=O)CN(CCN(CC(O)=O)CC(O)=O)CC(O)=O	industrial	Present in drinking water sources
Estrone	53-16-7	[H][C@@]12CCC(=O)[C@@]1(C)CC[C@]1([H])C3=C(CC[C@@]21[H])C=C(O)C=C3	natural origin	Baken et al. 2018
Ethyldimethylcarbamaat	687-48-9	CCOC(=O)N(C)C	industrial	Present in drinking water sources
exo-1,2,7,7-Tetramethylbicyclo[2.2.1]heptan-2-ol	2371-42-8	-	natural origin	Baken et al. 2018
F53B	73606-19-6	[K+].[O-]S(=O)(=O)C(F)(F)C(F)(F)OC(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)Cl	industrial	Present in drinking water sources
Fenoprofen	31879-05-7	CC(C(O)=O)C1=CC(OC2=CC=CC=C2)=CC=C1	pharmaceutical	Baken et al. 2018
Fexofenadine	83799-24-0	CC(C)(C(O)=O)C1=CC=C(C=C1)C(O)CCCN1CCC(CC1)C(O)(C1=CC=CC=C1)C1=CC=CC=C1	pharmaceutical	Present in drinking water sources
Flecaïnide	54143-55-4	FC(F)(F)COC1=CC(C(=O)NCC2CCCCN2)=C(OCC(F)(F)F)C=C1	pharmaceutical	Present in drinking water sources
Gadolinium compounds	-	-	contrast medium	Baken et al. 2018
Hordenine	539-15-1	CN(C)CCC1=CC=C(O)C=C1	natural origin	Present in drinking water sources
Hydroxyatrazine	2163-68-0	CCNC1=NC(O)=NC(NC(C)C)=N1	pesticide metabolite	Present in drinking water sources

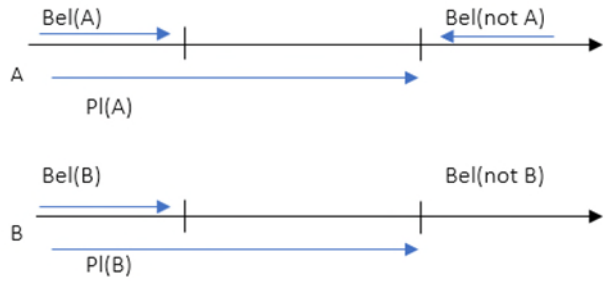
Name	CAS	SMILES	Origin	Reference
Hydroxysimazine	2599-11-3	CCNC1N=C(O)NC(NCC)=N1	pesticide metabolite	Present in drinking water sources
Iodopropinyl butylcarbamaat	55406-53-6	CCCCNC(=O)OCC#CI	fungicide	Present in drinking water sources
Ioxaglic acid	59017-64-0	CNC(=O)C1=C(I)C(N(C)C(C)=O)=C(I)C(C(=O)NCC(=O)NC2=C(I)C(C(=O)NCCO)=C(I)C(C(O)=O)=C2I)=C1I	contrast medium	Baken et al. 2018
Ioxitalamic acid	28179-44-4	CC(=O)NC1=C(I)C(C(=O)NCCO)=C(I)C(C(O)=O)=C1I	contrast medium	Baken et al. 2018
Ketamine	6740-88-1	CNC1(CCCCC1=O)C1=C(Cl)C=CC=C1	pharmaceutical	Present in drinking water sources
Levocetirizine	130018-77-8	OC(=O)COCCN1CCN(CC1)[C@H](C1=CC=CC=C1)C1=CC=C(Cl)C=C1	pharmaceutical	Present in drinking water sources
Levonorgestrel	797-63-7	[H][C@@]12CC[C@@](O)(C#C)[C@@]1(CC)CC[C@]1([H])[C@@]3([H])CCC(=O)C=C3CC[C@@]21[H]	pharmaceutical	Present in drinking water sources
Levopropylhexedrine	6192-97-8	-	pharmaceutical	Present in drinking water sources
Maleonitril	928-53-0	-	industrial	Present in drinking water sources
MDA	4764-17-4	CC(N)CC1=CC=C2OCOC2=C1	illicit drug	Present in drinking water sources
Melamine, hexa(methoxymethyl)-	68002-20-0	-	industrial	Baken et al. 2018
Methane, bromotrichloro-	75-62-7	ClC(Cl)(Cl)Br	industrial	Baken et al. 2018
Methocarbamol	532-03-6	COC1=CC=CC=C1OCC(O)COC(N)=O	pharmaceutical	Present in drinking water sources
Metolachlor ESA	171118-09-5	CCC1=CC=CC(C)=C1N(C(C)COC)C(=O)CS(O)(=O)=O	pesticide metabolite	Baken et al. 2018
Metolachlor OA	152019-73-3	CCC1=CC=CC(C)=C1N(C(C)COC)C(=O)C(O)=O	pesticide metabolite	Baken et al. 2018
MTDC	3013-02-3	CSC(=O)N(C)C	pharmaceutical	Baken et al. 2018

Name	CAS	SMILES	Origin	Reference
N-ethylperfluoroctaan-sulfonamidoazijnzuur	2991-50-6	CCN(CC(O)=O)S(=O)(=O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	industrial	Present in drinking water sources
Nicotine	54-11-5	CN1CCC[C@H]1C1=CN=CC=C1	natural origin	Present in drinking water sources
Octahydro tetramethyl naphthalenylethanone (OTNE)	54464-57-2	CC1CC2=C(CC1(C)C(C)=O)C(C)(C)CCC2	fragrance	Present in drinking water sources
Oxcarbazepine	28721-07-5	NC(=O)N1C2=CC=CC=C2CC(=O)C2=CC=CC=C12	pharmaceutical	Present in drinking water sources
Pyridoxine	65-23-6	CC1=C(O)C(CO)=C(CO)C=N1	pharmaceutical	Present in drinking water sources
Ritalinezuur	19395-41-6	OC(=O)C(C1CCCCN1)C1=CC=CC=C1	pharmaceutical metabolite	Present in drinking water sources
Sulfamic acid	5329-14-6	NS(O)(=O)=O	industrial	Present in drinking water sources
Sulfamide, N,N-dimethyl-	3984-14-3	CN(C)S(N)(=O)=O	pesticide metabolite	Baken et al. 2018
Sulpiride	15676-16-1	CCN1CCCC1CNC(=O)C1=C(OC)C=CC(=C1)S(N)(=O)=O	pharmaceutical	Present in drinking water sources
Tetraglyme	143-24-8 / 70992-84-6	COCCOCCOCCOCCOC	industrial	Present in drinking water sources
Traseolide	68140-48-7	CC(C)C1C(C)C(C)(C)C2=CC(C)=C(C=C12)C(C)=O	fragrance	Present in drinking water sources
Trifluoroacetic acid	76-05-1	OC(=O)C(F)(F)F	industrial	Present in drinking water sources
Triglyme	112-49-2 / 70992/85-7	COCCOCCOCCOC	industrial	Present in drinking water sources
Triphenylphosphine oxide	791-28-6	O=P(C1=CC=CC=C1)(C1=CC=CC=C1)C1=CC=CC=C1	industrial	Present in drinking water sources

Name	CAS	SMILES	Origin	Reference
Trospium chloride	10405-02-4	[Cl-].OC(C(=O)O[C@@H]1C[C@@H]2CC[C@H](C1)[N+]21CCCC1)(C1=CC=CC=C1)C1=CC=CC=C1	pharmaceutical	Present in drinking water sources
unknown	unknown	CCC1=C(C(=CC]C1)C)N(C(C)CO)C(=O)CCl	transformation product	Unpublished data KWR
Urea, N,N'-dicyclohexyl-	2387-23-7	O=C(NC1CCCCC1)NC1CCCCC1	pharmaceutical	Baken et al. 2018
Vigabatrin	60643-86-9	-	pharmaceutical	Present in drinking water sources

II Dempster Shafer Theory





Bel(A or B) = m1 + m2 + m3 + m6 + m9

m1	m4	m7
m2	m5	m8
m3	m6	m9

Pl(A or B) = 1 – m7

m1	m4	m7
m2	m5	m8
m3	m6	m9

Bel(A and B) = m3 = Bel(A) * Bel(B)

m1	m4	m7
m2	m5	m8
m3	m6	m9

Pl (A and B) = m2 + m3 + m5 + m6

m1	m4	m7
m2	m5	m8
m3	m6	m9

Bel(not A) = 1 – pl(A)

m1	m4	m7
m2	m5	m8
m3	m6	m9

Pl(not A) = 1 – Bel(A)

m1	m4	m7
m2	m5	m8
m3	m6	m9

Bel(not A and not B)= (1-Pl(A))*(1-Pl(B))

m1	m4	m7
m2	m5	m8
m3	m6	m9

Pl(not A and not B) = (1-Bel(A))*(1-Bel(B))

m1	m4	m7
m2	m5	m8
m3	m6	m9

Bel(not A or not B)= Bel(not A) + Bel(not B) = 1-Pl(A)+1-Pl(B)

m1	m4	m7
m2	m5	m8
m3	m6	m9

Pl(not A or not B)= Pl(not A) + Pl(not B) = 1-Bel(A)+1-Bel(B)

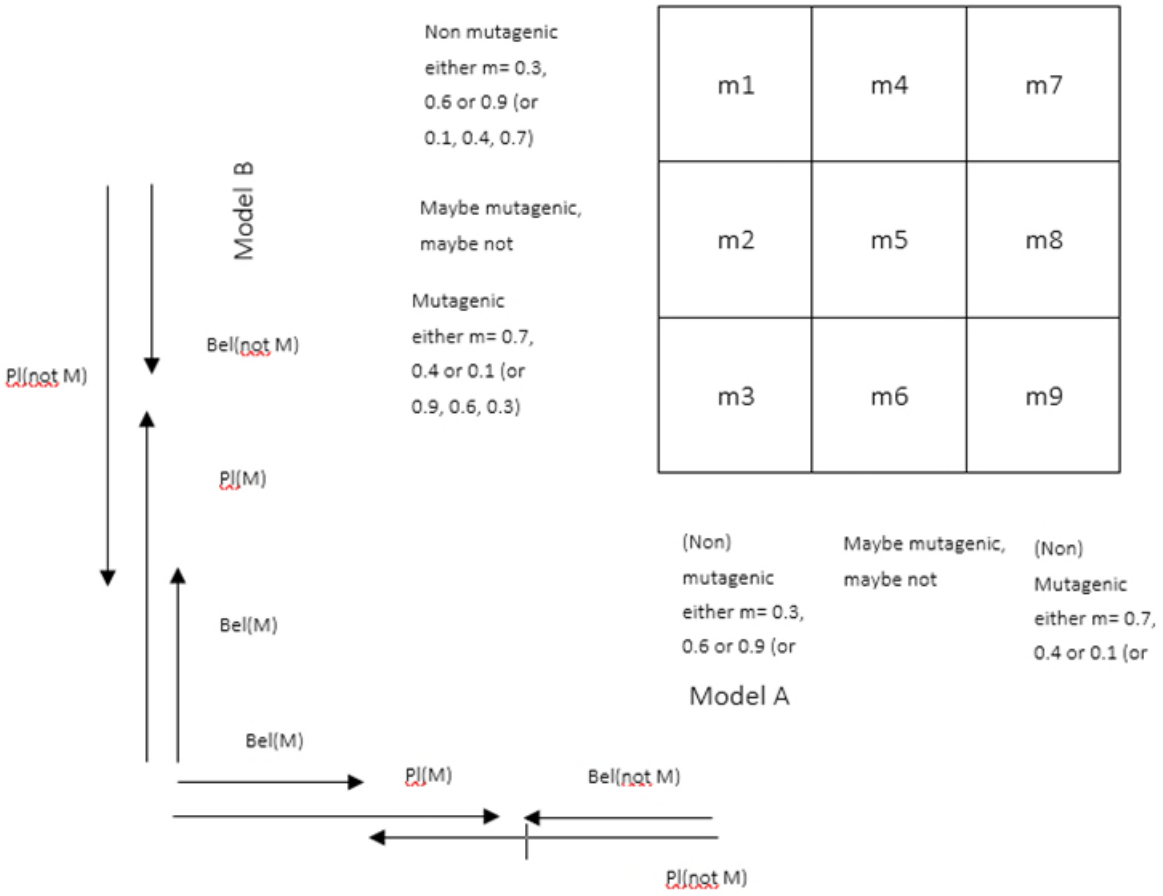
m1	m4	m7
m2	m5	m8
m3	m6	m9

The models in the VEGA QSAR use number of stars could indicate possibility.
Assuming 1 star = 0.3, 2 star = 0.6, 3 star = 0.9, although in reality this may be >0.3, >0.6 etc.
There can be multiple scenarios, we either:

- Consider Mutagenic/ possible mutagenic/ non mutagenic as categories (Bel(A), Pl(A), Bel(not A)), excluding information on the stars.
- Within the models, we assume 1 star 1-Pl(), 2 star Pl(), 3 star Bel(). However there are four models, so there are four dimensions, instead of two. Per model:

m1	m5	m9	m13
m2	m6	m10	m14
m3	m7	m11	m15
m4	m8	m12	m16

When a model reports that a compound is mutagenic (reliability for instance is 0.6), based on assumptions Bel(not mutagenic) = 0.1? Then this means that Bel(mutagenic) = 0.6, Pl(mutagenic) = 0.3+0.6.



Four models (A, B, C and D) with two categories so:

Chance that compound is mutagenic based on all models:

Bel(A and B and C and D) = Bel(A) * Bel(B) * Bel(C) * Bel(D) – Basically multiplying all probabilities

Change that compound is likely/maybe mutagenic based on all models:

Pl(A and B and C and D) = Pl(A)*Pl(B)* Pl(C)*Pl(D)

In a scenario in which only mutagenic results are reported this is 0.9*0.9*0.9*0.9