BTO 2018.030 | March 2018

BTO report

Tools for human health risk assessment of emerging chemicals

KWR Watercycle Research Institute

BTO

Tools for human health risk assessment of emerging chemicals

BTO 2018.030 | March 2018

Project number
400554/127/001

Project manager
Stefan Kools

Client
BTO - Thematical research - Emerging substances

Quality Assurance
Annemarie van Wezel

Author
Kirsten Baken

Sent to
This report is distributed to BTO-participants.
A year after publication it is public.

Year of publishing
2017

More information
Dr. Kirsten Baken
T 0031 30 6069 703
E kirsten.baken@kwrwater.nl

Keywords

Postbus 1072
3430 BB Nieuwegein
The Netherlands

T +31 (0)30 60 69 511
F +31 (0)30 60 61 165
E info@kwrwater.nl
I www.kwrwater.nl

KWR Watercycle Research Institute

BTO 2018.030 | March 2018 © KWR

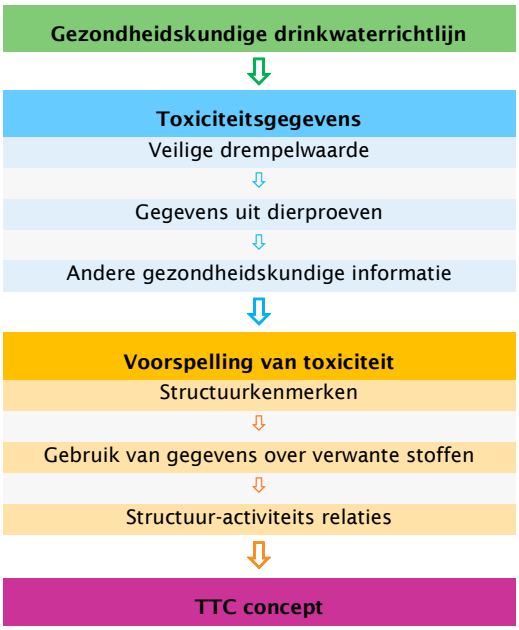
All rights reserved.
No part of this publication may be reproduced, stored in an automatic database, or transmitted, in any form or by any means, be it electronic, mechanical, by photocopying, recording, or in any other manner, without the prior written permission of the publisher.

BTO *Managementsamenvatting*

Selectie van informatiebronnen en 'in silico' tools bieden handvatten voor toxicologische risicobeoordeling van nieuwe stoffen

Auteur Dr. Kirsten Baken

Dit onderzoek biedt handvatten voor een snelle inschatting van potentiële gezondheidseffecten van stoffen wanneer klassieke toxicologische risicobeoordeling niet mogelijk is bij het aantreffen van nieuwe stoffen in drinkwater(bronnen). Bij zo'n klassieke aanpak worden stofconcentraties vergeleken met drempelwaarden die aangeven beneden welk niveau onze gezondheid zelfs bij levenslange blootstelling geen gevaar loopt. Echter, voor lang niet alle stoffen zijn gegevens beschikbaar om een drempelwaarde vast te stellen, met name in het geval van opkomende stoffen of transformatieproducten. Dan moeten analyses worden uitgevoerd waaruit een potentiële schadelijkheid kan worden afgeleid, waarbij de onbekende stoffen qua structuur worden vergeleken met verwante stoffen waarvan de toxicologische eigenschappen wel bekend zijn. Informatiebronnen en softwarematige tools – '*in silico*' tools – die hiervoor voorhanden zijn, zijn in dit onderzoek geselecteerd op basis van hun bruikbaarheid voor toxicologische evaluatie van drinkwatercontaminanten, gevolgd door een toelichting bij hun gebruik en toepassing voor praktijkvoorbeelden. Het resultaat is een workflow die helpt in het structureren van het gebruik van deze hulpmiddelen. De *in silico* tools geven niet specifiek een veilig blootstellingsniveau aan, maar zijn wel geschikt om stoffen snel te screenen op schadelijke eigenschappen en op basis daarvan te prioriteren voor verder onderzoek. Doorlopende inventarisatie van het aanbod van *in silico* tools en training in het gebruik ervan zijn nodig om de ontwikkelingen hierin bij te houden.



Workflow voor toxicologische evaluatie van chemische stoffen in drinkwater. Bij afwezigheid van informatie wordt overgegaan naar de volgende stap (van boven naar beneden).

Belang: bij gebrek aan toxiciteitsgegevens toch stoffen gezondheidskundig kunnen duiden

Voor het aanduiden van mogelijke gezondheidseffecten van chemische stoffen in drinkwater(bronnen) en het prioriteren van stoffen voor monitoring en zuivering is een toxicologische risicobeoordeling nodig. Zijn de beschikbare gegevens te onvolledig om zo'n risicobeoordeling uit te voeren, dan bieden *in silico* tools (informatiebronnen en software) een mogelijke uitkomst. Deze tools geven niet specifiek een veilig blootstellingsniveau aan, maar zijn wel geschikt om stoffen snel te screenen op schadelijke eigenschappen en op basis daarvan te prioriteren voor verder onderzoek.

Aanpak: inventarisatie, training en toetsing aan praktijkvoorbeelden

Voor een goede selectie en toepassing van *in silico* tools is specifieke expertise nodig, zo bleek uit eerder BTO-onderzoek. Ook moeten voor een volledige inschatting van toxicologische stofeigenschappen meerdere tools naast elkaar worden gebruikt. Vanuit deze behoefte is in dit onderzoek een inventarisatie uitgevoerd van

toxicologische informatiebronnen en bruikbare, publiek beschikbare *in silico* tools. Daarnaast is deelgenomen aan twee trainingen waarin experts ingingen op toepassingen van de tools. Dit rapport geeft een toelichting op de geselecteerde tools, inclusief *tips & tricks* voor het gebruik ervan. Tot slot laten we praktijkvoorbeelden zien waarin met de tools informatie wordt verkregen over stoffen waarvoor klassieke toxicologische risicobeoordeling tekortschiet.

In de OECD QSAR Toolbox zijn de grootste verzameling aan toxiciteitsgegevens en de meeste functionaliteiten en analysemogelijkheden beschikbaar. Aanvullende tools die geëvalueerd werden zijn AMBIT, ToxRead en T.E.S.T. voor het ophalen van toxiciteitsgegevens en voorspellen van toxiciteit; Toxtree, Chemotyper, ChemMine, Toxmatch, AIM, en VEGA voor het zoeken naar verwante chemische structuren en/of onderzoek naar toxicologisch eigenschappen; en Metaprint2D-REACT en Xenosite voor het voorspellen van metabolieten van chemische stoffen.

Resultaten: gestructureerde werkwijze voor toxicologische evaluatie van stoffen in drinkwater

De geselecteerde *in silico* tools zijn geïntegreerd in een workflow voor toxicologische evaluatie van chemische stoffen in drinkwater(bronnen). De workflow is gebaseerd op de informatie die men bij voorkeur bij evaluaties hanteert. Toepassing van de workflow op praktijkvoorbeelden laat zien dat *in silico* tools complementair zijn aan elkaar en richting kunnen geven aan de risicobeoordeling.

Implementatie: training en up-to-date kennis bevordert het gebruik van *in silico* tools

De workflow die uit dit onderzoek voortkomt, is toepasbaar voor de toxicologische evaluatie van nieuw aangetroffen stoffen in drinkwater(bronnen). Om waterbedrijven hierbij van dienst te zijn voorziet KWR in geregistreerde toxicologen die getraind zijn in het gebruik van de benodigde *in silico* tools, inclusief interpretatie van de resultaten. Gezien de snelle ontwikkelingen in ontwerp van software en *in silico* tools is een regelmatige update van de in dit rapport beschreven informatie aan te bevelen.

Rapport

Dit onderzoek is beschreven in rapport *Tools for human health risk assessment of emerging chemicals* (BTO 2018.030).
Hiermee is een vervolg gegeven aan de aanbevelingen uit *Innovative testing strategies and their relevance for evaluating chemical drinking water quality* (BTO 2014.009).

More information
Dr. Kirsten Baken
T 0031 30 6069 703
E kirsten.baken@kwrwater.nl

KWR
PO Box 1072
3430 BB Nieuwegein
The Netherlands

KWR Watercycle Research Institute

Contents

1	Introduction	3
2	Toxicological evaluation	4
2.1	Drinking water guidelines	4
2.2	Measured data	6
2.3	Predicted toxicity	7
2.4	TTC approach	10
3	*In silico* tools	13
3.1	OECD QSAR Toolbox	15
3.2	AMBIT	16
3.3	ToxRead	17
3.4	T.E.S.T.	18
3.5	AIM	19
3.6	Toxtree	20
3.7	Chemotyper	21
3.8	ChemMine	22
3.9	Toxmatch	23
3.10	Metaprint2D-REACT	24
3.11	Xenosite	25
3.12	EPI Suite™	26
3.13	VEGA	26
4	Applications	28
4.1	Chemicals without toxicity data	28
4.2	New perfluorinated compounds	30
4.3	Disinfection byproducts	31
4.4	Metabolites of plant protection products	33
5	Conclusions and final remarks	35
References		37
Attachment I		**39**
CAAT academy hands-on training		
Attachment II		**40**
CEFIC-LRI hands-on Training		
Attachment III		**41**
Application of *in silico* tools for an example chemical		
Attachment IV		**66**
Data collection sheet		

1 Introduction

Toxicological evaluation of chemical contaminants in (sources of) drinking water is required to identify potential human health risks and prioritise chemicals for monitoring and abatement. If toxicity data obtained from experimental animal studies (or in some cases *in vitro* studies) are incomplete or absent, non-testing (*in silico*) approaches can be applied to guide hazard evaluation. Such tools do not indicate safe exposure levels but are suitable for quick identification of potential hazards, prioritization of compounds for further testing, and to provide mechanistic information. We previously concluded that expertise is required to select non-testing tools and to perform and evaluate hazard predictions, and that multiple non-testing approaches should be combined to obtain the best prediction of toxicity (Baken and Kools, 2014). This report presents publicly available data sources and *in silico* tools and organizes them in a workflow for evaluation of potential human health hazards and/or health risks of chemical drinking water contaminants. Tips and tricks for the use of the non-testing tools are provided and illustrated by a number of examples of applications of these methods. In the final chapter, conclusions and final remarks on the use of the presented tools are summarized.

2 Toxicological evaluation

Potential human health risks of chemical contaminants can be identified by defining safe intake levels of chemicals and comparing those to measured or predicted (drinking) water concentrations (Schriks et al., 2010; Baken et al., 2018). For a limited number of chemicals known to appear in drinking water, health-based statutory drinking water standards are in place. For most chemicals that emerge in surface and groundwater, however, drinking water guideline levels have not yet been derived. In order to gather information on the potential toxicological properties of substances without health-based statutory drinking water standards and the potential human health risks upon exposure via drinking water, a workflow as depicted in Figure 2-1 is followed. In a tiered (step-by-step) approach, available information on toxicity is retrieved and/or toxicological properties of chemicals are predicted by non-testing strategies.

2.1 Drinking water guidelines

Legal drinking water standards and drinking water guideline values derived by renowned institutes for health protection (see Table 2-1 for a selection) correspond to safe drinking water concentrations, provided that they are based on human health risk assessment and not on organoleptic properties or technical achievability. Such standards and guideline values are based on lifelong daily consumption of a standard volume of water by an average individual in the general population. A proportion of the acceptable exposure level for a chemical is allocated to drinking water, taking into account exposure via other routes as well. Acceptable exposure levels for genotoxic substances represent maximum additional cancer risk levels at lifetime exposure. For Dutch drinking water quality standards, published guideline values can be recalculated to default values of 2 liters of drinking water consumption per day, an average body weight of 70 kg, 20% allocation of total exposure to drinking water (unless information on other exposure routes justifies a different allocation factor), and a maximum lifelong additional cancer risk level of 1 in 10^6. When multiple health based exposure thresholds are reported, either the most conservative value or the value that is most evidence-based can be adopted.

TABLE 2-1 SELECTON OF INFORMATION SOURCES FOR DRINKING WATER STANDARDS AND GUIDELINE VALUES

Legal standards	Drinkwaterbesluit / Drinkwaterregeling
	Regeling materialen en chemicaliën drink- en warm tapwatervoorziening
	EU Drinking Water Directive
	US E.P.A. National Primary Drinking Water Standards and Regulations
	Public Health Goals OEHHA California
	Guidelines for Canadian Drinking Water Quality
	Australian Drinking Water Guidelines
Guideline values	WHO Guidelines for drinking-water quality
	US E.P.A. National Primary Drinking Water Standards and Regulations
	USGS Health-based Screening Levels
	Rijksinstituut voor Volksgezondheid en Milieu (RIVM)

INFORMATION	EVALUATION	SOURCES AND TOOLS
DRINKING WATER GUIDELINE		
Legal standard Guideline value *(§2.1)*	Adopt or recalculate	Legislation *(Table 2-1)* Institutes *(Table 2-1)*
⇩		
MEASURED DATA		
Established TDI, ADI, RfD, DNEL , VSD *(§2.2.1)*	Calculate provisional drinking water guideline value	Meta-databases *(Table 2-2)* Institutes *(Table 2-2)*
⇧	⇩	
Established LO/NO(A)EL *(§2.2.1)*	Calculate acceptable daily intake and provisional drinking water guideline value	Meta-databases *(Table 2-2)* Institutes *(Table 2-2)* OECD QSAR Toolbox *(§3.1)* AMBIT *(§3.2)*
⇧	⇩	
Other types of information *(§2.2.3)*	Calculate NOAEL, acceptable daily intake and provisional drinking water guideline value	Meta-databases *(Table 2-3)* Institutes *(Table 2-3)*
⇩		
PREDICTED TOXICITY		
Structural alerts *(§2.3.1)*	Chemical profiling	OECD QSAR Toolbox *(§3.1)* AMBIT *(§3.2)* ToxRead *(§3.3)* Toxtree *(§3.6)* Chemotyper *(§3.7)* VEGA
⇩		
Read across *(§2.3.2)*	Identify suitable analogues Collect toxicity data Predict endpoint of interest	OECD QSAR Toolbox *(§3.1)* AMBIT *(§3.2)* ToxRead *(§3.3)* AIM *(§3.5)* ChemMine *(§3.8)* Toxmatch *(§3.9)* VEGA *(§3.13)* *and metabolism and physic-chemical properties*
⇩		
QSAR *(§2.3.3)*	Check applicability domain Predict endpoint of interest	OECD QSAR Toolbox *(§3.1)* AMBIT *(§3.2)* ToxRead *(§3.3)* T.E.S.T. *(§3.4)* VEGA *(§3.13)*
⇩		
TTC APPROACH		
TTC-based drinking water guideline level (§2.4)	Check exclusion categories Indications for genotoxicity? Carbamate or organophosphate? Cramer classification	See 'Structural alerts' T.E.S.T. *(§3.4)* *In vitro* bioassays

FIGURE 2-1. WORKFLOW FOR HUMAN HEALTH RISK EVALUATION OF EMERGING CHEMICALS. TOOLS ARE EXPLAINED IN CHAPTER 3. THE SEARCH FOR INFORMATION STARTS AT THE TOP LEVEL; WHEN INFORMATION IS LACKING DATABASES AND TOOLS PRESENTED AT A LOWER LEVEL ARE USED.

2.2 Measured data

When no health-based drinking water guideline values are available, provisional drinking water guideline values can be derived from either established acceptable daily intake levels or toxicity data obtained in experimental animal studies. Schriks et al. (2010) and Baken et al. (2015, 2018) have published a range of such guideline values for emerging contaminants detected in Dutch (sources of) drinking water.

2.2.1 Established TDI, ADI, RfD, DNEL, or VSD

Tolerable Daily Intake (TDI), Acceptable Daily Intake (ADI), Reference Dose (RfD), Derived No Effect Level (DNEL), or exposure levels corresponding to a specified additional life time cancer risk (Virtually Safe Dose; VSD), such as a 1 in 10^6 risk level, can be retrieved from toxicological (meta)databases and websites or reports published by renowned institutes for human health protection (see Table 2-2 for a selection). Provisional drinking water guideline values (pGLV) are calculated using different equations for threshold chemicals (a) and non-threshold (genotoxic) chemicals (b):

(a) pGLV (µg/L) = [TDI, ADI, RfD, or DNEL (µg/kg bw/day) x 70 kg body weight x 20% drinking water allocation] / 2 L drinking water consumption
(b) pGLV (µg/L) = [VSD x 70 kg body weight] / 2 L drinking water consumption

TABLE 2-2 SELECTION OF INFORMATION SOURCES FOR MEASURED TOXICITY DATA

Meta-databases	International Toxicity Estimates for Risk (ITER)
	FURETOX
	OECD eChemPortal
	TOXNET
Institutes	Rijksinstituut voor Volksgezondheid en Milieu (RIVM)
	European Food Safety Authority (EFSA)
	European Chemicals Agency (ECHA)
	EC Scientific Commissions
	EU Pesticides Database
	U.S. EPA (IRIS)
	U.S. EPA (Chemistry Dashboard)
	Agency for Toxic Substances and Disease Registry (ATSDR)
	U.S. EPA Human Health Benchmarks for Pesticides
	Pesticide Properties DataBase (PPDB)
	WHO International Programme on Chemical Safety (IPCS)
In silico tools*	OECD QSAR Toolbox
	AMBIT
	ToxRead
	T.E.S.T.

** In some in silico tools, toxicity data in underlying databases are searchable.*

2.2.2 Established LO/NO(A)EL

When acceptable daily intake levels have not been reported, No Observed (Adverse) Effect Levels (NO(A)Els), Lowest Observed (Adverse) Effect Levels (LO(A)ELs) or Benchmark dose (BMD) levels, usually derived from animal experiments applying chronic oral exposure, can be used to calculate acceptable daily intake levels for non-genotoxic substances. These data are preferably retrieved from the information sources in Table 2-2, since the quality of the toxicity studies has then been assessed by expert panels. Alternatively, results from toxicity studies published in peer reviewed literature can be used. Uncertainty factors to correct for inter- and intra-species variation, duration of exposure, and adequacy of the available toxicity data (a factor of 10 is often used by default for each aspect) are applied to reported NOAEL, LOAEL and BMD values to derive safe lifelong daily intake levels.

2.2.3 Other types of information

For pharmaceuticals, acceptable daily intake levels are often not published. The lowest therapeutic dose (LTD) can in this case be used as a NOAEL value, to which an uncertainty factor of 100 for inter- and intraspecies variation is applied to derive the acceptable daily intake. Table 2-3 shows information sources for therapeutic doses of pharmaceuticals.

In vitro bioassays are used as alternatives to animal experiments to predict toxicological endpoints (hazard identification). Such tests are useful to indicate the presence of substances with a certain biological activity in water samples, and can indicate mechanisms of action and (relative) potency for toxic effects (such as genotoxicity). However, results of *in vitro* studies can as yet not directly be translated to safe human exposure thresholds. More information on available databases and in vitro to in vivo extrapolation can be found in Baken and Dingemans (2017).

TABLE 2-3 SELECTION OF INFORMATION SOURCES FOR THERAPEUTIC DOSES

Databases	Farmacotherapeutisch Kompas
	Geneesmiddeleninformatiebank (CBG)
Institutes	European Medicines Agency (EMA)
	WHO Defined Daily Dose (DDD)

2.3 Predicted toxicity

When toxicity data are absent or incomplete, non-testing tools can be applied for toxicological evaluation. These computational methods are based on the principle that the activity of a chemical can be predicted from its molecular structure and substructure(s), and from the physicochemical properties and biological effects of similar substances. A non-comprehensive overview of familiar and well characterised structures is provided by the US EPA (URL). Recent developments in computing power, the ability to create extensive databases and the use of the internet to compile, organise and distribute information, have increased the capability and capacity to investigate relationships between chemical structure and biological activity (Baken and Kools, 2014). *In silico* tools do not generate dose-response information required for human health risk assessment and thus cannot replace toxicity testing. *In silico* approaches can however be used additionally to results from experimental animal studies to increase the confidence in the available toxicity data. In absence of toxicity data these tools can indicate which threshold of toxicological concern (TTC) level would be appropriate (i.e. for genotoxic or non-genotoxic chemicals), and guide experimental approaches to gather toxicity data for data-poor chemicals by indicating which endpoints would be most relevant to assess primarily. A selection of publicly available *in silico* tools is presented in chapter 3. Types of information on potential toxicity that can be generated by such tools is described below.

2.3.1 Structural alerts

Structural alerts (SA) are functional groups or structural features that are qualitatively linked to the presence or absence of a property or activity. SA may be complete molecules or parts of molecules. Structure Activity Relationships (SARs) are based on knowledge of chemicals with known physicochemical properties and/or biological actions (reactions with biological entities such as cells or molecules) (Raies and Bajic, 2016; Bower et al, 2017). When a SA is identified in a chemical structure, the description of the SA should be checked for human relevance and for reliability. The reactivity of a SA can be modified by other elements present in the chemical structure; subtle differences in chemical structures and the position of the SA in the molecule may therefore have a considerable impact on the biological activity. Insight

in modulating substructures is however limited for most SA. SA for genotoxicity have been studied most extensively (Kolkman et al., 2013). SA are used to identify potential chemical hazards, group compounds into categories for read across (§2.3.2), predict toxicity in QSAR models (§2.3.3), and assign chemicals to TTC categories (§2.4).

2.3.2 Read across

Chemical categorizing refers to the process of grouping of chemicals whose properties and/or fate are likely to be similar as a result of structural similarity. Read across techniques can subsequently be applied to fill data gaps on a specific target chemical by interpolating or extrapolating existing data of related chemicals within a category (Raies and Bajic, 2016; Bower et al, 2017). In principle, analogue-based read across can be applied for any property or endpoint, irrespective of whether it is a physicochemical property, environmental fate parameter, human health effect, or ecotoxicological effect (Baken and Kools, 2014), provided that (i) the profile of the chemical of interest is known, (ii) suitable analogues are identified, and (iii) measured toxicity data of adequate quality (e.g. as indicated by Klimisch score) are available for the analogues (Schultz et al., 2015). OECD and US EPA have categorized chemicals within the OECD HPV Chemicals Programme or a category defined within the chemical notification scheme and the HPVC challenge programme of the US EPA, and a target chemical may thus already have been associated with an existing category. The OECD QSAR Toolbox (§3.1) and AIM (§3.5) informs the user when this is the case.

It is advised to use multiple tools to identify analogues in order to cover as many chemical databases as possible. Analogues can be identified and selected based on various characteristics (Figure 2-2), that are all equally important. *In silico* tools that allow identification of analogues may offer options to set thresholds for structural similarity scores, indicate critical structural alerts, simulate metabolism, and retrieve analogues with measured toxicity data only. When (dis)similarity in chemical structure is noted, it is important to check in which part of the molecule the (dis)similarity is located and whether that particular part is:

- present in both the target and the analogue,
- critical for the chemical properties,
- representing a functional group or SA,
- related to the endpoint of interest.

Relevant physico-chemical properties are melting point, boiling point, vapor pressure, water solubility and log Kow (Petry, 2017; Autiero, 2017). When only part of the analogues have measured data for these parameters, it is advised to use predicted physico-chemical properties for all substances for optimal comparability (Autiero, 2017).

Taking both structural and biological similarity into account will increase the quality of the category. Kinetic behavior of a chemical in an organism related to absorption, distribution, metabolism and excretion (ADME) properties of chemicals affects the toxicity. Parameters that determine ADME properties are log P, log D, pKa, water solubility, bioavailability, membrane permeability, availability of transporters, volume of distribution, plasma protein binding, blood-brain barrier penetration, metabolic transformation, and excretion. *In silico* models generally predict these properties with moderate accuracy, due to limited availability of data to train the models. Prediction of the structures and subsequently the toxicity of metabolites yields a more complete toxicological evaluation of chemicals and comparison between targets and analogues. Simulation of metabolism requires the consideration of parameters involved in absorption and distribution (which determine the concentration at the target site), interaction with metabolic enzymes, and prediction of the structure and reactivity of all metabolites that are formed. *In silico* tools generally only assess one of those aspects. In addition, metabolic pathways are complex due to the multitude of enzymes that may be involved, differences between species, individuals, organs and tissues, and influence

of other internal and external factors such as age, disease, and stress. For cytochrome P450 (CYP450) metabolism and prediction of metabolite structures, several *in silico* models are available. However, most models cannot discriminate between major and minor or stable and reactive metabolites, and produce many false positives. Therefore, results of multiple *in silico* models should be compared and collated to identify the primary metabolites (Kirchmair, 2017).

Characteristic s of chemicals	**Similarity**			
• Structural features & functional groups	Similar	Similar	Major substructure shared	Not shared
• Structural alerts & reactivity	Similar	Similar	Key reactive groups shared	Not shared
• Effect of other molecular features on toxicity	No	No	No	Yes
• Physico-chemical properties and toxicokinetics	Similar	Not similar but no effect on toxicity		Dissimilarity alters toxicity
• Metabolic pathways	Similar	Similar	Overlap in parents/ metabolites	Dissimilar activity of metabolites
Decision: suitable analogues for read across?	**Suitable**	**Suitable with interpretation**	**Suitable with preconditions**	**Not suitable**

FIGURE 2-2 DECISION SCHEME FOR IDENTIFICATION OF SUITABLE ANALOGUES FOR READ ACROSS ANALYSIS (BASED ON AUTIERO, 2017 AND SCHULTZ ET AL., 2015)

The decision which analogues and read across approaches are suitable depends on expert judgement. Read across for hazard assessment can be qualitative or quantitative. In qualitative read-across, the presence (or absence) of a property/activity for the target chemical is inferred from the presence (or absence) of the same property/activity for one or more analogues. In quantitative read-across, the known value(s) of a property for one or more source chemicals is used to estimate the unknown value of the same property for the target chemical. In the case of a toxicological effect, this approach implies that the potency of an effect shared by the two chemicals is similar or follows a regular pattern (Baken and Kools, 2014).

ECHA has published practical guidance documents on how to use and report (Q)SARs in the REACH registration process (ECHA, 2008, 2016, 2017). ECHA discriminates between 'analogue' and 'category' approaches (Figure 2-3). In the analogue approach, read across is based on a very limited number of structurally similar substances, where no trend or regular pattern in the properties is apparent. The term category approach is used when read across is employed between several substances that have structural similarity and as a result of this similarity, the toxicological and/or environmental fate properties are expected to be similar or follow a regular pattern.

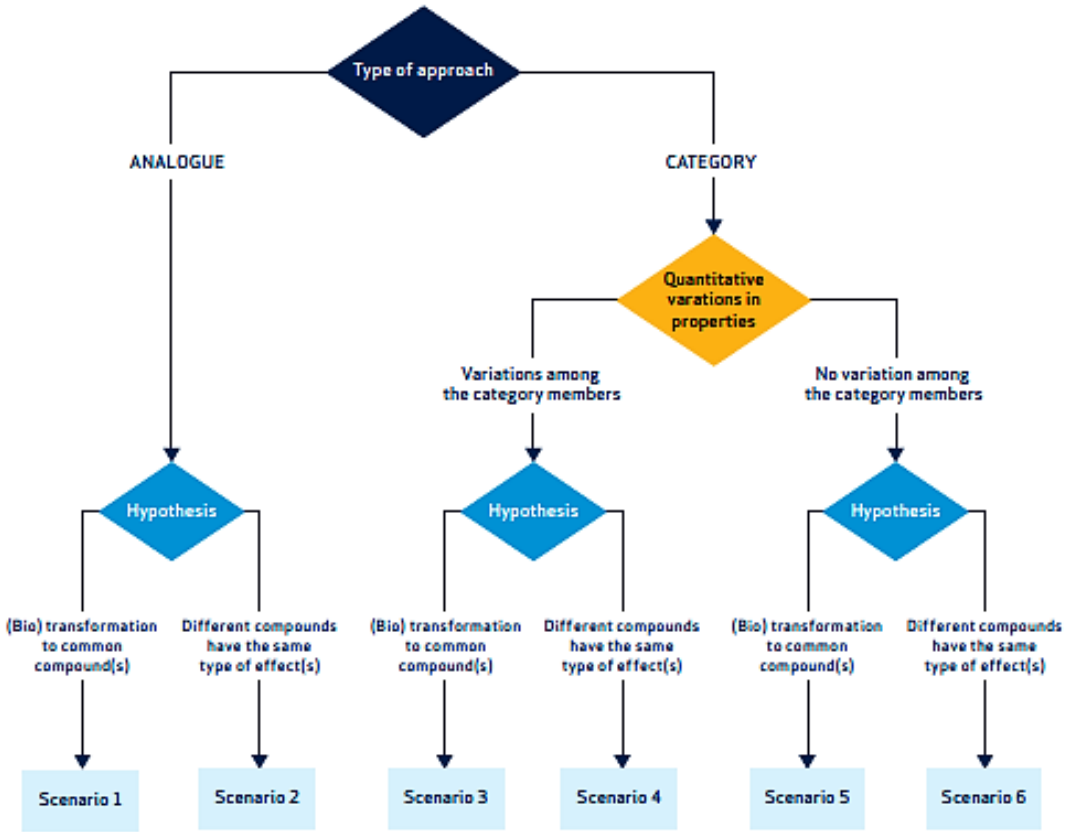


FIGURE 2-3 SCENARIOS IN THE ECHA READ ACROSS ASSESSMENT FRAMEWORK (RAAF), REFLECTING DIFFERENT TYPES OF READ ACROSS APPROACHES (ECHA 2012A)

2.3.3 QSAR

Quantitative Structure Activity Relationships (QSARs) are mathematical models that quantitatively predict the physicochemical, biological (e.g. toxicological) and environmental fate properties of molecules based on knowledge of the chemical structure. QSARs are more prevalent for endpoints for which large databases exist, such as ecotoxicity, mutagenicity and carcinogenicity, skin sensitisation, and endocrine disruption. For complex endpoints (such as repeated dose toxicity and reproductive and developmental toxicity) the models are not overly realistic because they may ignore essential processes (Raies and Bajic, 2016; Bower et al, 2017). Baken and Kools (2014) have provided an overview of available QSAR models. The EC Joint Research Center (JRC) offers an up-to-date QSAR Model Database that provides information on the validity of QSAR models that have been submitted to the JRC (URL). A selection of publicly available *in silico* tools in which QSAR models are embedded are presented in chapter 3.

2.4 TTC approach

The TTC is a pragmatic approach, providing conservative exposure limits based on information on chemical structure in absence of toxicity data. The concept originates from the Threshold of Regulation (ToR) that was based on carcinogenicity data for hundreds of chemicals (Rulis, 1986). TTC levels have been calculated for groups of non-genotoxic chemicals (i.e. Cramer class I, II and III, referring to presumed degree of systemic toxicity) based on No Observed Adverse Effect (NO(A)EL) values of reference substances derived from animal experiments (oral dosing) on (sub)chronic, reproductive and developmental toxicity (Munro et al. 1996). A separate threshold for certain neurotoxicants and pesticides (i.e.

organophosphates and carbamates) and a TTC threshold level specifically designed for carcinogens with a structural alert for genotoxicity have been published (Kroes et al. 2004).

TTC-based drinking water target values have been derived from TTC levels by Mons et al. (2013). The generic TTC level for carcinogenic substances and the TTC for genotoxic substances were translated to drinking water concentrations using the approach described in §2.2.1, providing conservative threshold levels of 0.1 and 0.01 µg/L for non-genotoxic and genotoxic compounds, respectively. In a recent evaluation, we evaluated these TTC-based drinking water target values levels based on toxicity data of substances detected in drinking water and sources, and concluded that somewhat higher thresholds may be used for non-genotoxic chemicals (Baken and Sjerps, 2016; Baken et al. 2018). Table 2-4 lists the different TTCs and TTC-based drinking water target values.

TABLE 2-4 TTC LEVELS AND TTC-BASED DRINKING WATER GUIDELINE VALUES

Classification	TTC (µg/day)	Reference	TTC-based drinking water target value (µg/L)	Reference
Cramer class I (*low toxicity*)	1800	*Munro et al. 1996*	37.7	*Baken and Sjerps, 2016*
Cramer class II (*medium toxicity*)	540	*Munro et al. 1996*		
Cramer class III (*high toxicity*)	90	*Munro et al. 1996*	4.0	*Baken and Sjerps, 2016*
Organophophates and carbamates	18	*Kroes et al. 2004*		
Carcinogens	1.5	*TOR rule ('80)*	0.1	*Mons et al. 2013*
Genotoxic substances (except aflatoxins, azoxy- or N-nitroso compounds)	0.15	*Kroes et al. 2004*	0.01 0.02	*Mons et al. 2013; Baken and Sjerps, 2016*

The TTC approach should not be applied to substances with complex chemical structures having multiple structural elements and highly unique structures, such as some pharmaceuticals (SCCS, 2012). Other substances that are excluded from the TTC approach, either due to underrepresentation in the databases or because they may still be of toxicological concern at the TTC exposure levels, include high potency carcinogens (i.e. aflatoxin-like, azoxy- or N-nitroso-compounds, benzidines, hydrazines), inorganic substances, metals and organometallics, proteins, steroids, organosilicon compounds, chemicals that are known or predicted to bioaccumulate, nanomaterials, radioactive substances, and mixtures of substances containing unknown chemical structures (Kroes et al. 2004; EFSA, 2012; EFSA/WHO, 2016). Such substances need to be evaluated on a case by case basis by gathering experimental (genotoxicity) data.

When substances do not belong to the exclusion categories and can be assigned to the chemical classes for which TTC values have been defined, TTC-based drinking water target values can be used as safe exposure levels for chemicals with unknown toxicity present in drinking water and its sources. First, potential genotoxicity needs to be identified for both the chemicals of interest and its (predicted) metabolites. Chemicals that are classified as genotoxic can be retrieved from the information sources indicated in Table 2-5. Indications for genotoxicity can be derived from *in vivo* or *in vitro* tests assessing gene mutations (e.g. the Ames test) or chromosomal aberrations (including micronuclei) (Table 2-2 and 2-5), structural alerts or read across. DNA reactivity (including direct interaction of chemicals with DNA and covalent modification of DNA) is of most concern, since for other genotoxicity mechanisms safe exposure threshold can often be derived (Kroes et al. 2004; EFSA/WHO,

2016; Boobis et al. 2017). OECD and OASIS DNA binding profilers are structural alerts included in a wide range of chemicals that are mainly designed for grouping and are considered over-predictive (i.e. yielding false positive alerts). Such alerts should preferably be combined with QSAR predictions (Boobis et al. 2017).

TABLE 2-5 SELECTION OF INFORMATION SOURCES ON GENOTOXICITY

Organisations	IARC
	NTP
	U.S. EPA / IRIS
	ECHA
	RIVM
	Ministerie van Sociale Zaken en Werkgelegenheid
Databases	TOXNET GENOTOX
	TOXNET CCRIS
	ECVAM

Non-genotoxic chemicals can be assigned to Cramer classes. The Toxtree tool (§3.6) is often used to automatically go through the Cramer decision tree. An extended decision tree is included for more accurate classification, at least for compounds consisting of one benzene ring. Since Cramer class II contains few chemicals, the TTC threshold for Cramer class III, which is possibly more robust since it is based on a more representative number of compounds, may conservatively be used for chemicals categorized in class II as well (EFSA/WHO, 2016).

3 *In silico* tools

In silico tools assess the (toxicological) properties of chemicals using computer-based estimations or simulations. While human health risk assessment based on reported toxicity data is common practice for emerging drinking water contaminants, *in silico* tools are not yet routinely applied in this area. Here, we present freely available tools for human health hazard assessment, in particular with respect to the oral exposure route and chronic toxicity, that can be used in context of the workflow presented in Figure 2-1. These tools were selected based on earlier evaluations and recommendations by expert users (Baken and Kools, 2014; Boobis et al. 2017; CAAT Academy Hands-on training, see Attachment I). Table 3-1 shows some specifications of the tools and the steps in the workflow (Figure 2-1) in which they can be applied. The tools are further explained below. Attachment III illustrates the information that can be retrieved by each tool taking the chemical 5-methyl-1H-benzotriazole, an emerging chemical in the water cycle, as an example.

In most tools, the chemical of interest can be indicated by name, CAS number, and/or SMILES code. It is advised to verify CAS and SMILES codes using multiple information sources such as the OECD QSAR Toolbox (see below), ChemIDplus (https://chem.nlm.nih.gov/chemidplus), U.S. EPA Chemistry Dashboard (https://comptox.epa.gov/dashboard), and PubChem (https://pubchem.ncbi.nlm.nih.gov). SMILES codes can be saved in a file with .sdf extension using for instance https://cactus.nci.nih.gov/translate or http://chemmine.ucr.edu.

For all tools it should be noted that reliable predictions can only be made when the chemical of interest is within the applicability domain of the embedded model regarding the physicochemical, structural, or biological properties, knowledge or information of the training set on which the model or tool is based. If this is the case, the target chemical is well represented by the training set. Salts, polymers and surfactants are for instance often not part of the training set.

Attachment IV contains a data collection sheet that can be used to integrate the information retrieved using the *in silico* tools. This aids in identification of data gaps, interpretation of data collected, and drawing conclusions based on the 'weight of evidence' by the user. When conflicting results are obtained for a specific endpoint, either the prediction with the highest confidence level or the most conservative prediction can be used.

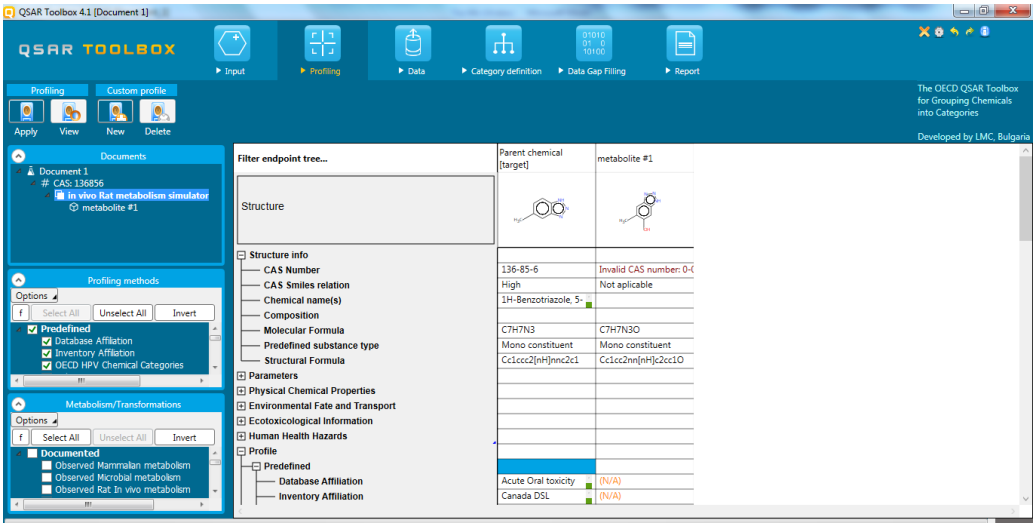
TABLE 3-1 CHARACTERISTICS AND FUNCTIONALITIES (SEE EXPLANATION IN §2.3) OF A SELECTION OF PUBLICLY AVAILABLE IN SILICO TOOLS FOR TOXICOLOGICAL EVALUATION OF CHEMICALS. THE ORDER OF THE FUNCTIONALITIES (HORIZONTALLY) AND TOOLS (VERTICALLY) FOLLOW THE LEVELS IN THE WORKFLOW IN FIGURE 2-1.

Tool	Online / Download	Input	Measured data	SA (genotoxicity)	SA (other)	Structural analogues	Physico-chemical properties	Observed metabolites	Metabolism simulator	Read across	QSAR (toxicological endpoints)	Exclusion category	Cramer class	Batch mode	Output	Remarks
OECD QSAR Toolbox	D	Name CAS SMILES SMART Structure	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓ (Toxtree)	✓ (only test data and profiling)	• (Selection of) data matrix in Excel • Customized assessment report	Largest number of databases and structural alerts included
AMBIT2	O*,D	Name CAS SMILES Structure	✓	✓ (Toxtree)	✓	✓	✓	~ *(when present as substance constituent)*	-	✓	✓ (VEGA)	-	✓ (Toxtree)	–	• Excel file of data or working matrix • Word report of assessment • Sdf and other file types of structures	**Use Chrome browser* All constituents of substances are displayed.
ToxRead	D	SMILES or sdf	✓	✓	-	✓	✓ (LogP)	-	-	✓ (Ames)	✓	-	-	-	Mutagenicity yes/no	Limited selection of endpoints
T.E.S.T.	D	CAS SMILES Structure	✓	-	-	-	✓	-	-	-	✓ (Ames)	-	-	✓	• Conclusio n of Ames test • Prediction score	Limited selection of endpoints
AIM	D	Name CAS SMILES Structure	~ *(links only)*	-	-	✓	-	-	-	-	-	-	-	-	Document with analogues and hyperlinks to data sources	No detailed report
Toxtree	D	CAS SMILES Structure	-	✓	✓	-	-	-	✓ (SMARTCyp)	-	-	✓	✓	✓	Decision tree result and most probable CYP450 metabolites	Stand alone version runs only one prediction at a time
Chemotyper	D	sdf	-	✓	-	-	-	-	-	-	-	-	-	✓	Highlighted structural alerts	
ChemMine	O*	SMILES or sdf	-	-	-	✓	-	-	-	-	-	-	-	✓	Structures and similarity scores; download SMILES or sdf	**Use Chrome or Firefox browser*
ToxMatch	D	SMILES or sdf	-	-	-	✓*	-	-	-	-	-	-	-	✓	Similarity graph and score; export to Excel	**no search, only similarity check*
Metaprint2D -REACT	O	SMILES Structure	-	-	-	-	-	-	✓ (phase I and II)	-	-	-	-	-	Structures and SMILES of analogues	No export or SMILES codes of results available
Xenosite	O	SMILES or sdf	-	~ *(DNA reactivity)*	-	-	-	-	✓ (sites of metabolism)	-	-	-	-	-	Pictures or sdf file	Predicts one reaction at a time
Epi Suite	D	Name CAS SMILES Structure	-	~ *(physchem properties)*	-	-	✓	-	-	-	-	-	-	✓	Word or txt file	
VEGA	D	SMILES or sdf	✓	✓	✓		✓	-	-	✓	✓	-	-	✓	PDF or CSV	Also embedded in AMBIT & ToxRead

3.1 OECD QSAR Toolbox

Developer	Latest version	Website
LMC (Bulgaria)	4.1 (August 2017)	http://www.oecd.org/chemicalsafety/oecd-qsar-toolbox.htm

The OECD QSAR Toolbox supports the practical application of grouping of chemicals and read-across approaches for data gap filling. The Toolbox incorporates information and tools from various sources and is regularly updated under peer review of OECD member state countries, ECHA, chemical industry and NGO's. Compared to previous versions, the current version contains updated profilers and metabolic simulators, increased ADME information, new databases, a streamlined workflow, and a reliability score for alerts and databases. The results are presented in a transparent way, linking to further information and underlying data.



The OECD QSAR Toolbox can be applied to perform various assessments:

- Collection of measured data for a single or multiple chemicals[1] from 49 databases with 2 million data points retrieved from animal and *in vitro* studies. Links to the databases from which the data originate are provided. When specific endpoints are selected, a colour code indicates which databases are considered suitable.
- Identification of relevant structural characteristics and potential mechanism of action of one target chemical or multiple structures at the same time[1] and/or their observed or predicted metabolites (which appear at the bottom of the endpoint tree). Chemical profiling can be done based on pre-defined, general mechanistic, endpoint specific, empiric, toxicological, and/or metabolic characteristics. When specific characteristics are selected, a colour code indicates which profilers are suitable. Right clicking on a profiler shows an explanation (some are for instance 'under development') and the applicability domain. Clicking on a SA flagged for a chemical in its profile shows the decision tree and the part of the molecule that forms the SA. When the chemical of interest is out of domain for a prediction or profiler, this is indicated.
- Retrieve observed (mammalian, rat, rat liver, rat S9 or microbial) metabolites and predict metabolites (autoxidation, dissociation, hydrolysis, rat, rat S9, skin, microbial or tautomerism). The predicted metabolites are ranked based on probability and can

[1] To collect data for multiple chemicals simultaneously, right click on the first chemical and choose 'Add in category'.

be shown along with the target in the working matrix[2], after which they can each be profiled.

- Identification of analogues using various criteria. 'Structure similarity' is the broadest profiler (under 'Options', choose the 'Dice' index and 'Atom centred fragments' for similarity calculation with a threshold of 70% to start with), which can be followed by a subcategorization based on 'organic functional groups' or other (endpoint specific) characteristics to narrow down the number of hits; a colour code again indicates suitable characteristics. All predicted metabolites of the analogues can be retrieved at the same time ('Define with metabolism'). Additional analogues indicated by other tools can be added manually (right click on target -> 'Add in category'). The overlap in structural alerts, physico-chemical properties, bioavailability (Lipinski rule), and metabolism and the availability of measured data can be checked after profiling all analogues (and metabolites) and gathering measured data from selected databases. The structural similarity of all analogues can be revealed by right clicking the target structure next to 'Parameter'->'2D'.
- Perform read across or trend analysis (in case of sufficient continuous quantitative data that are expected to show a trend) for data gap filling for a selected endpoint. The chemical category can be refined/adapted during the procedure; the effect on the prediction is visualised instantly. When multiple measurement results are available for an analogue, the option 'Data usage' -> 'Maximum' allows a worst case prediction.
- Run embedded QSARs for physico-chemical properties, environmental fate, Developmental and Reproductive Toxicity (DART), irritation, sensitization and ecotoxicity

3.2 AMBIT

Developer	Latest version	Website
Ideaconsult Ltd (Bulgaria)	AMBIT2 v3.1.0 (September 2017)	https://ambitlri.ideaconsult.net/tool2

AMBIT has been developed within the CEFIC-LRI research program to support category approaches and read across and establishing a valid justification. The tool is regularly updated and extended; future plans include design of AMBIT as a central hub that connects different tools, including EPI Suite (§3.12). AMBIT2 contains updated non-confidential REACH data provided by ECHA as well as the new European Food Safety Authority (EFSA) OpenFoodTox database (>450.000 substances in total). It discriminates between 'substances' and the constituents of which substances are composed ('structures', additives, metabolites and impurities). The category formation and data gap filling procedure is structured by a workflow. Assessment reports can be generated automatically.

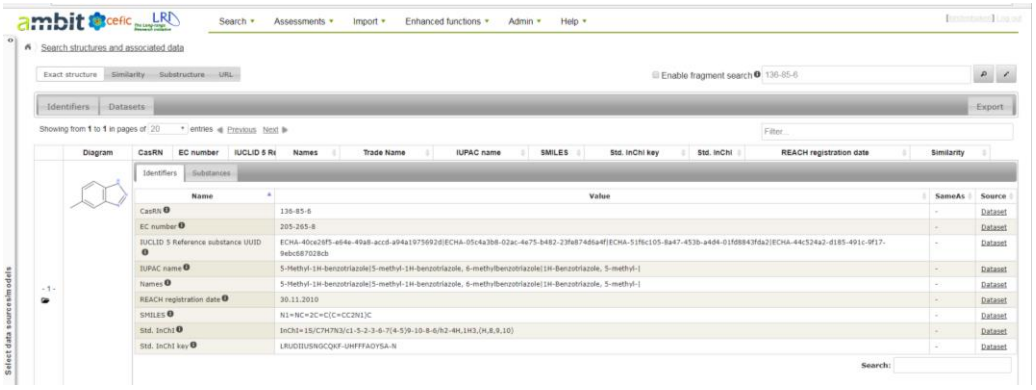
The following functions are available in AMBIT2 (Jeliaskova, 2017; see Attachment II):

- A 'Search' mode for substances (by name) and related constituents, or structures (by name, CAS or SMILES) and substances containing this structure as a constituent (the folder icon displayed for each search result discloses this information).
- Measured data can be retrieved for 43 endpoints. Toxicity data are reported for substances, not structures. Relatively little high quality (i.e. Klimisch score 1 or 2) ECHA data are available. Users need to select data for read across manually by either specifying 'Reliability' in 'Advanced search' or deleting data from the working matrix afterwards.
- Via 'Enhanced functions' Toxtree toxicity predictions (structural alerts, see §3.6) and the VEGA model (QSAR, see §3.13) can be run. For VEGA predictions, the ADI score

[2] To display all metabolites along with the target, right click on the chemical identifier in the left menu of the Input page -> 'Multiplication'.

(referring to applicability domain) indicates the reliability (1 = experimental data). To include Toxtree and VEGA predictions in the data matrix, 'Datasets/Models' need to be checked in the vertical sidebar at the left of the screen.

- The 'Assessment' mode allows category formation and read across. Analogues of structures (only by SMILES code) are identified by structural similarity ('Tanimoto similarity'). A similarity score is indicated and a threshold can be set. When data are gathered for analogues, users need to check whether the identified analogue is a mono-constituent of the substance for which data are reported, or whether a substance contains the analogue as an impurity or metabolite. Physicochemical properties and measured toxicity data for all substances related to each analogue can be unfolded. Next, the target structure (T) and category members (CM) need to be selected; a rationale can be added manually. Endpoint data to be used for read across are selected in the next step. 'Supporting information' in the 'Tox' tab refers to model predictions (Toxtree and VEGA). An initial data matrix is generated automatically by the selection of substances and endpoints. Data points can be added or removed in the working matrix. Data gap filling results need to be derived and added by the user.



3.3 ToxRead

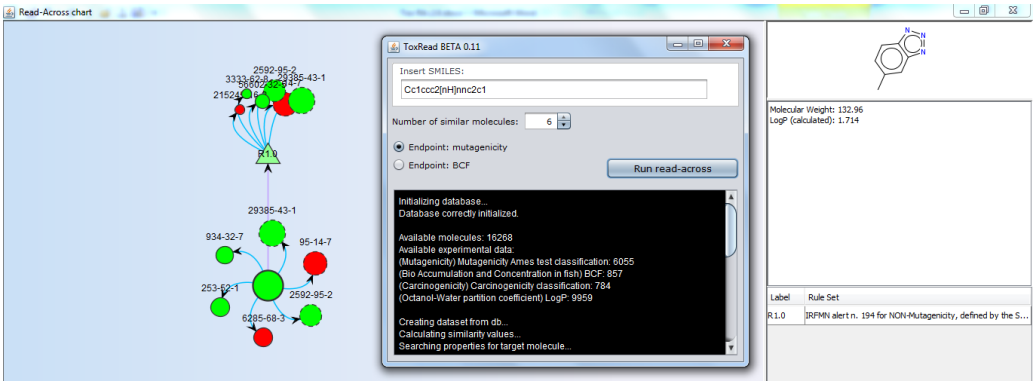
Developer	Latest version	Website
Mario Negri (Italy)	0.11 (September 2016)	http://www.toxread.eu

ToxRead assists users in making reproducible read across evaluations. The current version contains a read across mode for mutagenicity (Ames test) and bioconcentration. Other endpoints are under development. The user has to provide the chemical of interest, the endpoint, and the number of similar chemicals to be included. Similar chemicals are automatically selected by the software. ToxRead contains:

- Libraries of chemicals with associated experimental values. In the database, 16268 molecules are included with experimental values for Ames test classification (6055 data points), carcinogenicity (784 data points), bioaccumulation and -concentration in fish (857 data points), and octanol-water partition coefficient (LogP: 9959 data points). These data originate from LIFE projects ANTARES, CALEIDOS and PROSIL.
- Libraries of structural alerts and algorithms of relevant features. These were derived from the aforementioned projects and from VEGA (§3.13) and Toxtree (§3.6) libraries. For the mutagenicity endpoint, four rule sets are available: Benigni/Bossa (as available in Toxtree v. 1), SARpy rules (extracted by Politecnico di Milano, with the automatic tool SARpy), IRFMN rules (extracted by human experts at Istituto di Ricerche Farmacologiche Mario Negri) and CRS4 rules (extracted by CRS4 Institute with automatic tools).

Results are displayed in graphs, in which the target chemical is connected with:

- The most similar compounds. These are represented by circles, the size of which is proportional to the similarity, ranging from 0 to 1. The user can see the structure of the chemical of interest by double clicking on the circle. ToxRead shows the structure, the similarity values, and the experimental values for a series of endpoints, not only for the endpoint under direct evaluation.
- Structural alerts. These are represented by triangles, which are inverted for non-toxic alerts. Toxicity alerts are red, while non-toxic alerts are green. The intensity of the colour is related to the percentage of toxic or non-toxic chemicals. Clicking on the alert shows the chemical structure, the prevalence of toxic compounds, the P value associated to the alert (which is represented by the triangle size), and up to 100 chemicals most similar to the target chemical containing the alert. The user should prefer alerts with a higher P value and prevalence of toxic/non-toxic chemicals. In the graph, structural alerts are also connected with related chemicals.



For the mutagenicity endpoint, an integrated prediction is provided and on its basis the target molecule is depicted in green or red (mutagenic or non-mutagenic prediction), or in yellow if the prediction is conflicting. This prediction is based on two assessments:

i. an automatic calculation of a read across prediction based on the output of ToxRead. Molecules having an experimental value in disagreement with the toxicity reported by the alert are excluded. Two scores are calculated for mutagenic and non-mutagenic activities: each score is the sum of the predicted activity multiplied by the similarity for each compound having the same toxicity activity. The final scores are then normalized (in the range between 0 and 1) and the score with the highest value represents the read-across prediction.

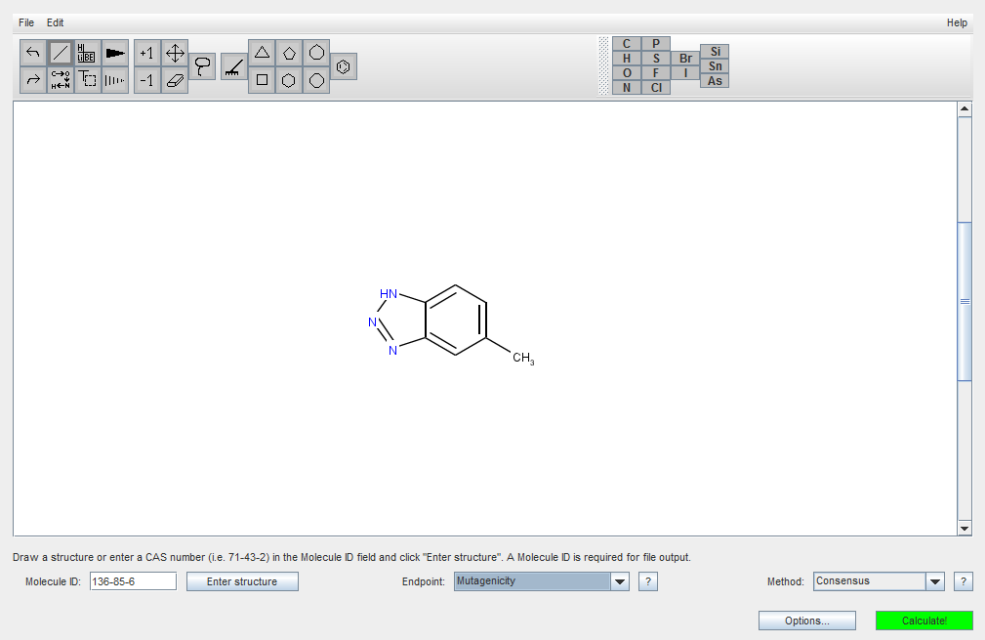
ii. a consensus (indicated by a Consensus Score) between four different QSAR models available in the VEGA software (see §3.13), and experimental data when available.

3.4 T.E.S.T.

Developer	Latest version	Website
US EPA	4.2.1 (April 2016)	https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test

The Toxicity Estimation Software Tool (T.E.S.T.) is an open-source application developed by the US EPA's National Risk Management Research Laboratory. T.E.S.T allows users to estimate a selection of toxicological and a range of physical properties. The toxicological endpoints that are currently in the software include ecotoxicity parameters, acute toxicity (rat oral LD50), developmental toxicity, and mutagenicity in the Ames test. When experimental results are present in the underlying database, these are reported as well. The physical property endpoints include boiling point, flash point, surface tension, viscosity, density,

water solubility, thermal conductivity, vapour pressure, and melting point. No predictions can be made for organic salts.

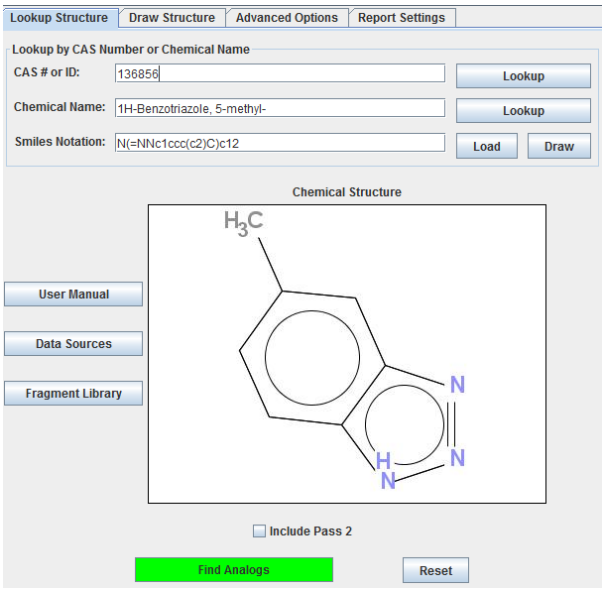


T.E.S.T. uses several QSAR methodologies: hierarchical, FDA, single-model, group contribution, nearest neighbour and consensus, and mode of action methods. The consensus method is preferred, since it predicts the toxicity simply by taking an average of the predicted toxicities from the aforementioned methods. To check the reliability of the predictions coming from this consensus method, the user should assess the reliability of the other methods. Predictions for similar structures are reported as well. If the predicted value matches the experimental values for similar chemicals in the training set, and the similar chemicals were predicted well, one can have greater confidence in the predicted value.

3.5 AIM

Developer	Latest version	Website
US EPA	1.01 (November 2013)	https://www.epa.gov/tsca-screening-tools/analog-identification-methodology-aim-tool

The Analog Identification Methodology (AIM) facilitates analogue analysis and data identification in support of chemical assessment or read across approaches. AIM conducts a structural analysis of a chemical using over 700 individual atoms, groups and super fragments indexed in a predefined database. It then matches them to potential analogues from a built in inventory of over 86.000 chemicals with publicly available measured data. ‘Pass 1’ is the default stringent search in which all of the fragments/atoms in the query chemical are contained in the analogue. ‘Pass 2’ can be selected by the user to include a less stringent search. Under ‘Advanced options’, additional rules can be selected to loosen search criteria even more, in case Pass 1 and 2 yield few analogues. AIM searches may also be tailored by the user to define what types of substitutions or exclusion rules are appropriate for the search. Polymers cannot be run in AIM. For ring structures, only exact matches can be performed.



A report is produced listing the target chemical and identified analogues. Similarity scores and overlapping structural characteristics are not reported. Hyperlinks to data sources with experimental results are provided; however, users still need to look up the specific chemical in the databases themselves. 21 common chemical classes known to undergo metabolism in the body to potential metabolites of concern for various health effects are flagged. The user needs to determine when a specific analogue is suitable for a specific assessment, as the determination of what structure is 'appropriate' can vary depending on the endpoint assessed.

US EPA also offers the Chemical Assessment Clustering Engine (ChemACE) tool (https://www.epa.gov/tsca-screening-tools/chemical-assessment-clustering-engine-chemace), which is designed to cluster a list of chemicals based on structure using predefined similarity rules. The ChemACE methodology uses the same fragment generation system found in AIM, but applies a more complex method for identifying analogs for the clustering exercises.

3.6 Toxtree

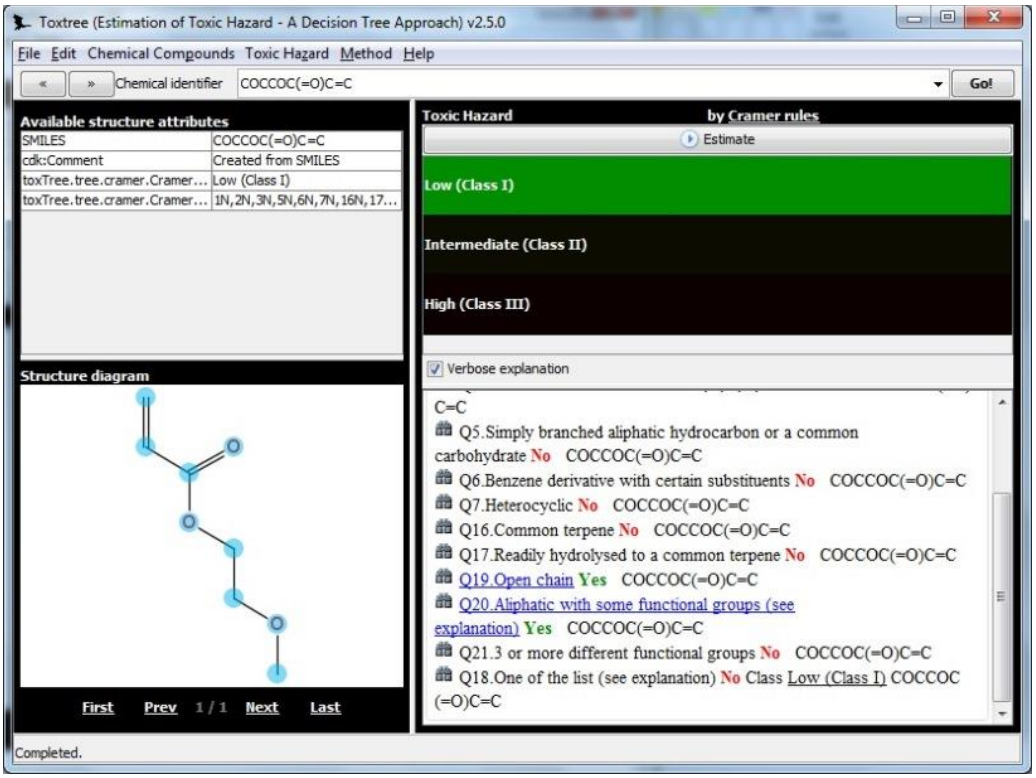
Developer	Latest version	Website
Ideaconsult Ltd (Bulgaria)	2.6.13 (March 2015)	http://toxtree.sourceforge.net/

Toxtree encodes a number of rulebases (sets of interrelated logical rules) for the evaluation of toxicity. It is an expert system of SARs that can be useful to identify potential hazards but also to provide the mechanistic information to substantiate read-across. Toxtree was originally commissioned by the JRC to encode the Cramer structural classes that are routinely used as part of the TTC approach. Since then, Toxtree has been extended and further developed with other rulebases.

Examples of its functionalities are:

- Decision tree for the application of the Threshold of Toxicological Concern (TTC) approach;
- Cramer scheme and an extended Cramer scheme;
- Mutagenicity and carcinogenicity rulebase known as the Benigni-Bossa rulebase as well as the ToxMic rulebase on the *in vivo* micronucleus assay;

- Rules to predict skin and eye irritation and corrosion and skin sensitisation;
- SMARTCyp, which predicts which sites in a molecule are targets for metabolism by Cytochromes P450. Good accuracy but limited coverage of reaction types and atom environments (Kirchmair, 2017). The metabolic reactions are ranked based on probability. Clicking on a result shows the site of metabolism and structure of the predicted metabolite, which can be copied and used as input for profiling or prediction of a subsequent metabolic step;
- START biodegradability, a set of structural alerts compiled by the Canadian EPA for estimating the biodegradability potential of a chemical compound based on structural alerts.



Some of the rulebases have been implemented or re-encoded into the OECD QSAR Toolbox (§3.1) and AMBIT (§3.2). The predictions may deviate from the original tool due to small differences in programming. These tools do however allow online use of ToxTree (AMBIT) and application of multiple rulebases at the same time, which is not possible in Toxtree itself.

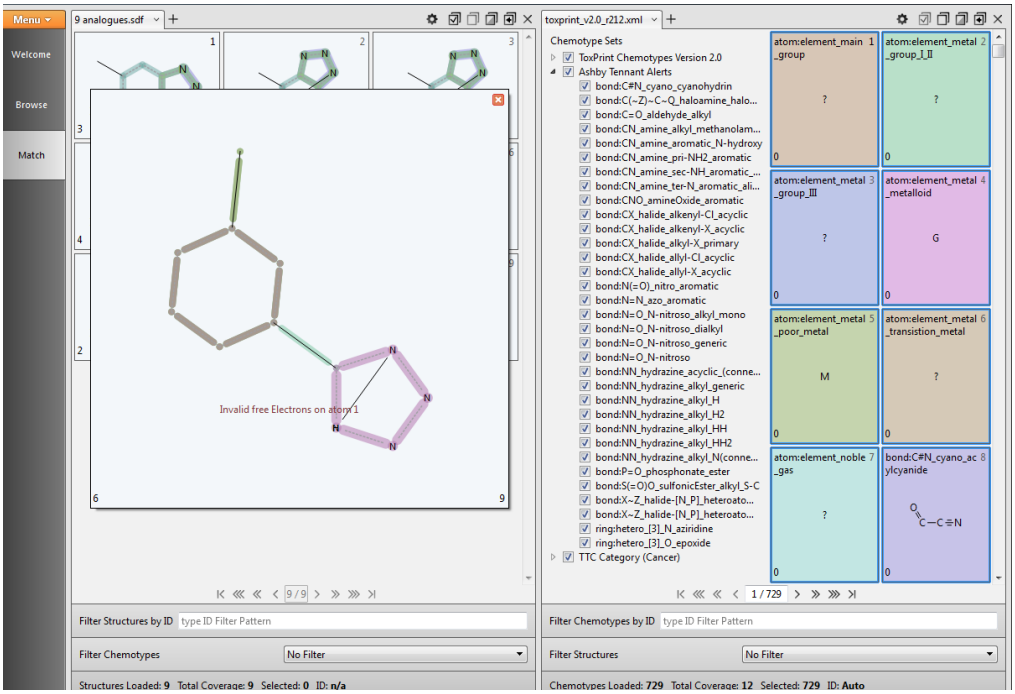
3.7 Chemotyper

Developer	Latest version	Website
Molecular Networks GmbH	1.0 (November 2013)	https://chemotyper.org

Chemotyper was developed under contract from FDA to house the public set of 'ToxPrint' chemotypes (chemical substructures or subgraphs), which were developed for FDA's CERES project. ToxPrint consists of over 700 individual chemotypes and contains the following three basic subsets:

- Generic structural fragments;
- Ashby-Tennant genotoxic carcinogen rules;
- Carcinogenicity alerts for TTC categorisation.

Chemotyper allows for searching and visualisation of the chemotypes and grouping of chemicals according to chemotypes. After loading the ToxPrint chemotypes (First option in the Welcome page), they can be matched to the target chemical(s). Chemotypes selected by the user are highlighted in the molecule.

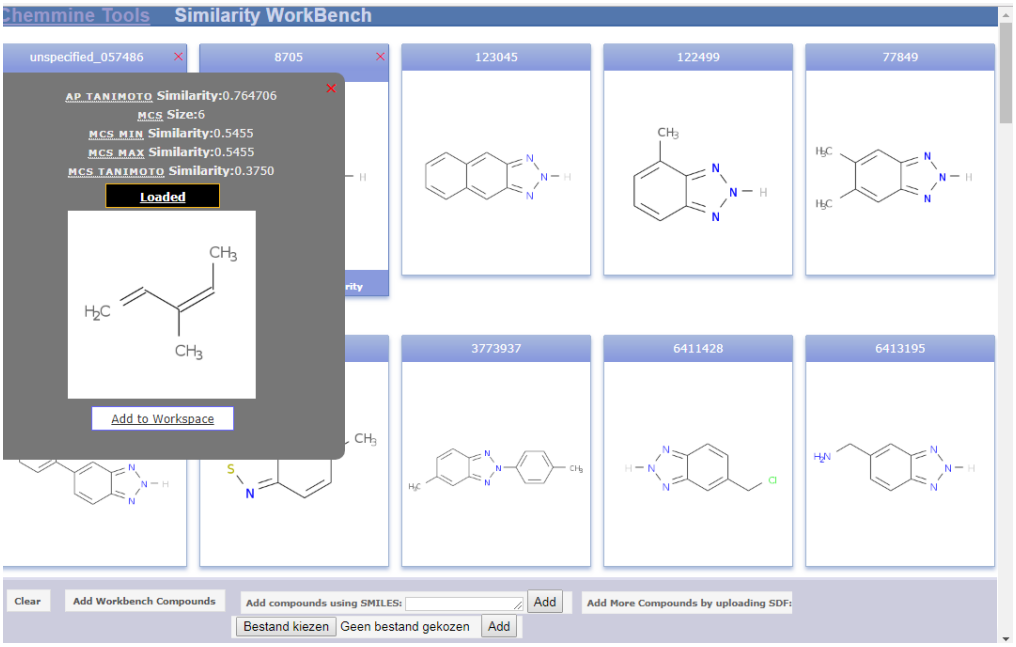


3.8 ChemMine

Developer	Latest version	Website
Girke Lab (USA)	2011	http://chemmine.ucr.edu

ChemMine has been developed for analyzing and clustering small molecules by structural similarities, physicochemical properties or custom data types. It can be used to find structural analogues. ChemMine is linked to PubChem, which provides the advantages of including recently developed chemicals in the search as well.

After adding the target chemical, the Option 'Search similar compounds' can be selected. A predefined similarity cutoff can then be chosen; it is advised to first start with a high cutoff level (0.9) and always use the 'Fingerprint algorithm'. A list with Hits appears, which can be sent to the Similarity Workbench. The analogues can each be compared with the target molecule visually and by a score for similarity (that appears when the target and an analogue are selected).



3.9 Toxmatch

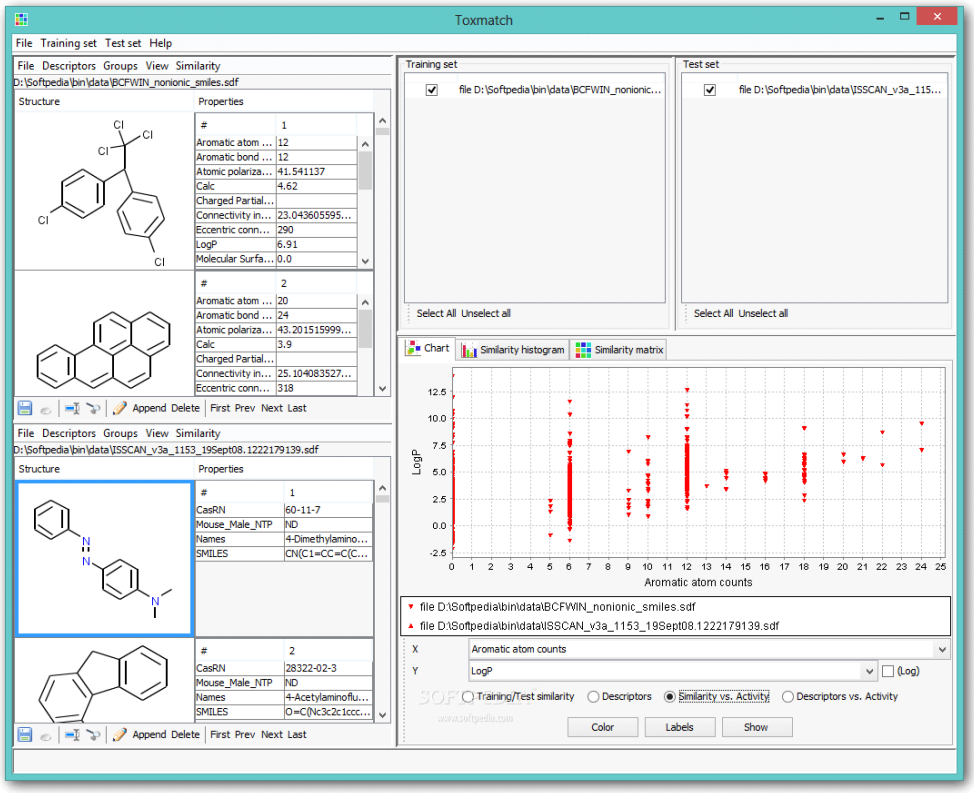
Developer	Latest version	Website
Ideaconsult Ltd (Bulgaria)	1.07 (January 2009)	https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/toxmatch

Toxmatch was developed under the terms of a JRC contract. The main functionalities of this advanced tool are to compare a query chemical to a training set in order to classify the chemical and to compare datasets, based on various structural and descriptor-based similarity indices. This aids in categorisation of chemicals for read across purposes. Similarity to the training set should preferably be assessed based on structural characteristics relevant to the endpoint. Predefined training sets for aquatic toxicity, bioconcentration factor, skin sensitization, skin irritation, carcinogenicity and mutagenicity are available. The carcinogenicity dataset originates from ISSCAN and contains 1153 chemicals with information pertaining to carcinogenicity (field 'Canc' with values 3: carcinogen, 2: equivocal and 1:non-carcinogen) and mutagenicity in Salmonella typhimurium (Ames test) (field 'SAL' with values 3:mutagen, 2: equivocal and 1: non-mutagen).

Users need to proceed through the following steps:

- Open training set in top left panel and select associated groups (e.g. mutagen/equivocal/non-mutagen);
- Open test set (single target chemical or group) in bottom left panel;
- Calculate or load and explore descriptors (chemical characteristics) for both data sets;
- Select 'Similarity to training set' in the top panel, use for instance 'Euclidean distance' or 'Tanimoto distance (fingerprints, kNN)' and finish with 'select Calculate similarity and predict activity';
- Choose View Fields from the top panel to select the properties used to calculate similarity in the similarity tab. Switch to descriptors panel and select all descriptors.
- Now run the similarity calculation for the test compound, following the previous two steps as above but now in the bottom panel.

Results are presented in various graphical displays including scatter plots, pair wise/ composite similarity histograms and similarity matrices. The graphs can be customized and similarity thresholds can be applied.



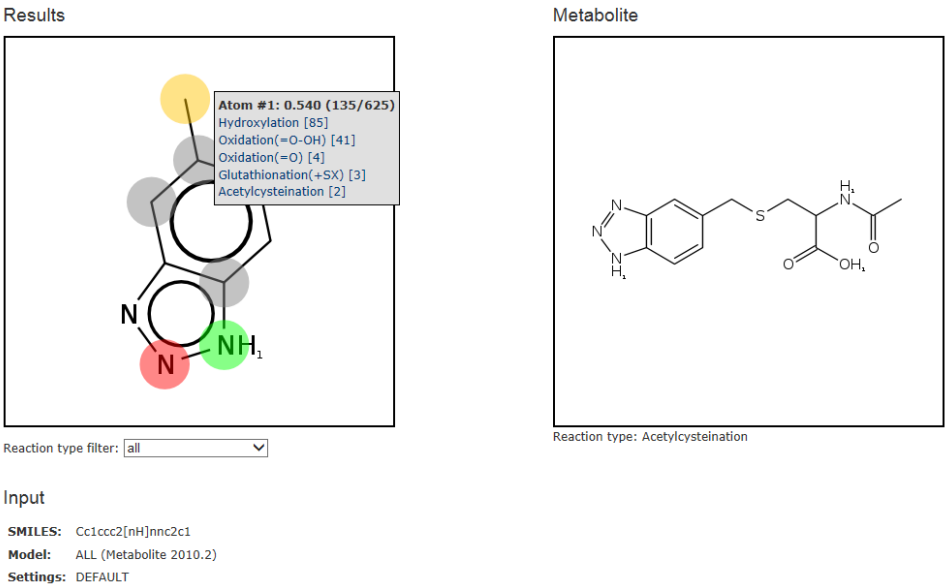
3.10 Metaprint2D-REACT

Developer	Latest version	Website
University of Cambridge	2007	http://www-metaprint2d.ch.cam.ac.uk

The web application of Metaprint2D-REACT was disabled in February 2018; it is currently not known whether the tool will become available again.

Metaprint2D derives the likelihood of metabolic transformations by mining large biotransformation databases including phase I and phase II metabolic pathways. The models assumes that each chemical has at least one target site for metabolism. The sites of metabolism are indicated in the chemical structure; clicking on an atom displays the predicted reactions at that site, and clicking on one of these reaction types shows the metabolites formed. The 'occurrence ratio' is shown for each site of metabolism, indicating the proportion of the instances of this specific atom in the database that is involved in the predicted metabolic transformation. This is commonly regarded as a rather simplistic way to predict metabolism (Kirchmair, 2017).

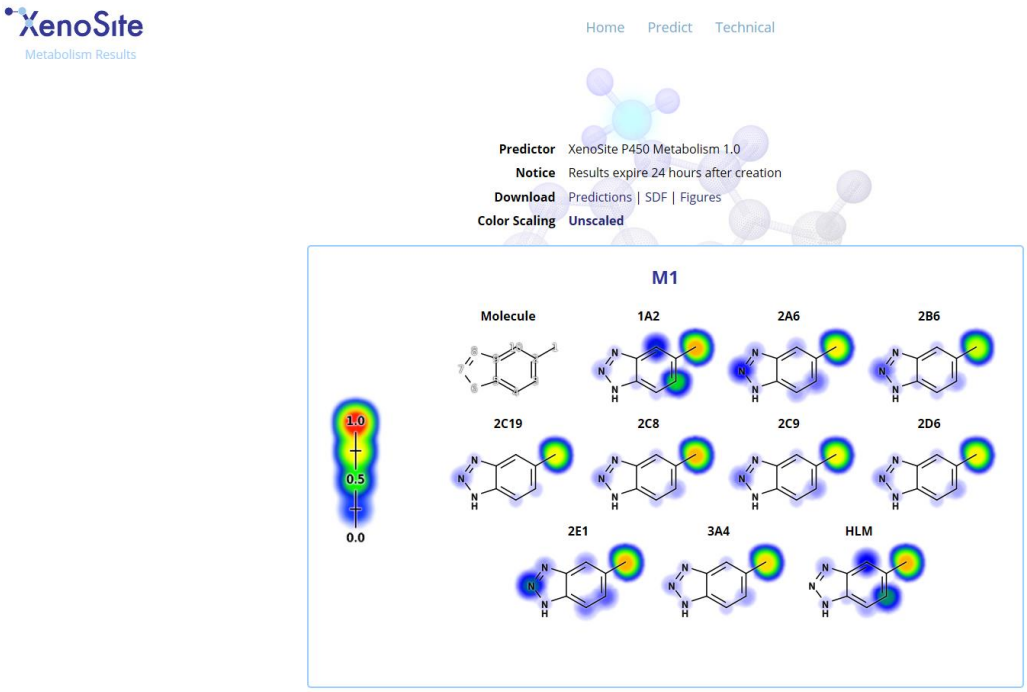
Metaprint2D-REACT generates structures of likely metabolites. The 'occurrence ratio' is shown between brackets, which can be regarded as the probability of this metabolic transformation. The tool particularly works well if a very comparable molecule is present in the database. Therefore, the more data that are used the better the prediction, and thus it is advised to model dog, human, and rat metabolism at the same time (Kirchmair, 2017).



3.11 Xenosite

Developer	Latest version	Website
Washington University	2017	http://swami.wustl.edu/xenosite

XenoSite is a tool for predicting the atomic sites at which chemicals are expected to undergo metabolic modification. It includes simulation of cytochrome P450 metabolism based on computations for which among others the SmartCyp software is used. Other types of reaction such as oxidation, reduction, and epoxidation are included too, as well as reactivity towards glutathione, DNA, and protein. Structures of potential metabolites are not provided.

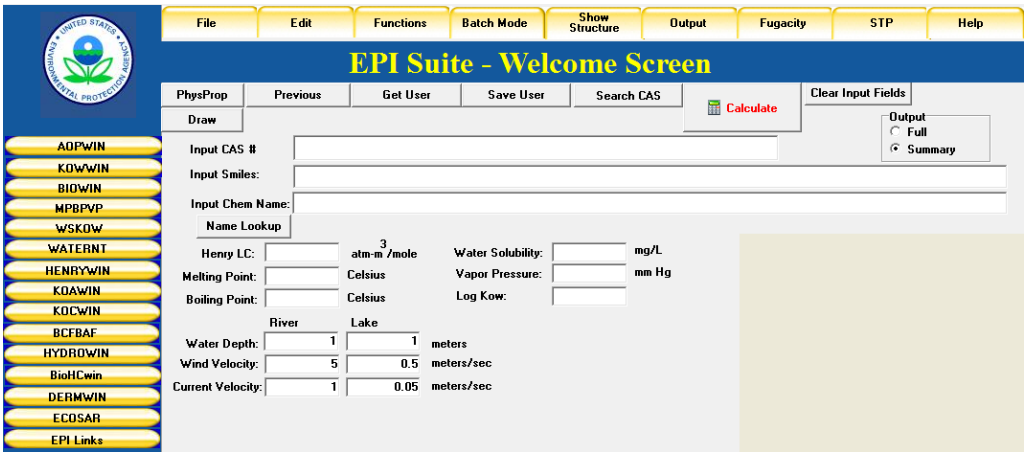


XenoSite provides visual output for each molecule and each enzyme. Potential sites of metabolism are labeled by a color gradient that indicates probability. The underlying data also show whether the observed site of metabolism is known from literature or used in the training set, and the background probability of observing a site of metabolism given the model, which can be used to interpret the prediction.

3.12 EPI Suite™

Developer	Latest version	Website
U.S. EPA	4.11 (November 2012)	https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface

EPI Suite (Estimation Programs Interface Suite) estimates a range of physicochemical properties, environmental fate parameters and ecotoxicity. It has been developed by the US Environmental Protection Agency (EPA) in collaboration with Syracuse Research Corporation (SRC). The applicability domain is based on molecular weight or LogKow. EPI Suite can be used for organic chemicals and organic salts with simple counter ions. It allows data gap filling for melting point, boiling point, vapour pressure, water solubility and octanol/water partition coefficient. EPI Suite is a screening-level tool and should not be used if acceptable measured values are available (except when not all analogues under evaluation have measured data, in which case predicted physico-chemical properties for all substances should be used for optimal comparability).



3.13 VEGA

Developer	Latest version	Website
Mario Negri (Italy)	1.1.4 (February 2017)	https://www.vegahub.eu/portfolio-item/vega-qsar

VEGA (Virtual models for property Evaluation of chemicals within a Global Architecture) is a QSAR model that was based on the CAESAR, T.E.S.T., SARpy, EPISuite, Toxtree, and other tools. It was designed to generate transparent results and is regularly updated. VEGA is also incorporated in AMBIT2 (§3.12) and Toxtree (§3.6). The applicability domain is checked by the program; no predictions can be made for organic salts. Although VEGA uses an advanced algorithm to calculate similarity of analogues (which is indicated by a score between 0 and 1; a score >0.75 should be aspired), the final evaluation on similarity should be done by the user. VEGA predicts e.g. Ames mutagenicity, carcinogenicity, developmental toxicity by several models (it is advised to combine the results of multiple models) and hepatotoxicity. The reliability of each prediction is indicated by the Applicability Domain Index.

VEGA Mutagenicity (Ames test) model (CAESAR) 2.1.13 page 1

1. Prediction Summary

Prediction for compound Molecule 0

Prediction: Reliability:

Prediction is Mutagenic, but the result shows some critical aspects, which require to be checked:
- only moderately similar compounds with known experimental value in the training set have been found
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 infrequent fragments found)
The following relevant fragments have been found: SA13 Hydrazine; SA27 Nitro aromatic

Compound: Molecule 0
Compound SMILES: O=[N+]([O-])c1oc(cc1)c2cnc(NN=C(C)C)s2
Experimental value: -
Predicted Mutagen activity: Mutagenic
Structural alerts: SA13 Hydrazine; SA27 Nitro aromatic
Reliability: the predicted compound could be out of the Applicability Domain of the model
Remarks:
none

It is also possible to use VEGA solely for read across. For this more limited use, the predicted value should be disregarded and instead the similar compounds need to be identified on the basis of similarities in mechanisms and descriptors.

4 Applications

The *in silico* tools described in the previous chapter are useful to gain insight in the potential human health hazard of chemicals present in drinking water or sources with incomplete toxicological information. Examples of such chemicals are substances of emerging concern, newly designed perfluorinated compounds, disinfection byproducts, and metabolites of plant protection products. The workflow presented in Figure 2-1 was applied to evaluate an example for each of these groups of chemicals. Results were documented in the data collection sheet (Attachment IV) and are presented below.

4.1 Chemicals without toxicity data

Chemicals that appear in the water cycle because they have only recently become in use and/or have not been emitted before, or because chemical-analytical methods were not able to detect them earlier, often lack a complete toxicological dossier. In previous BTO research, for 21 out of 163 currently detected contaminants in drinking water (sources) no provisional drinking water guideline value could be derived because of absence of chronic toxicity data (Baken et al. 2018). One of those chemicals is 5-methyl-1H-benzotriazole. Results of its evaluation using the *in silico* tools presented in chapter 3 are presented in Attachment III; the data collection sheet is shown below.

	Target chemical	Analogues	
Name	5-methyl-1H-Benzotriazole	1,2,3-Benzotriazole	Tolytriazole
CAS	136-85-6	95-14-7	29385-43-1
SMILES	Cc1ccc2[nH]nnc2c1	C1=CC2=NNN=C2C=C1	CC1=CC=CC2=NNN=C12
Metabolites			
Observed (O) / Predicted (P)	No observed metabolites	No observed metabolites.	No observed metabolites.
CYP450	4 metabolites predicted by Toxtree: OCc1ccc2NN=Nc2(c1) is the most probable according to MetaPrint and Xenosite); no measured data, genotoxicity and reprotox alerts present	Comparable sites of metabolism; no overlapping metabolites	Comparable sites of metabolism; one similar metabolite
rat S9	4 metabolites predicted by OECD Toolbox (3 are very probable according to MetaPrint and Xenosite); no measured data, genotoxicity alerts present	2	0
rat *in vivo*	1 metabolite overlaps with rat S9. MetaPrint does not consider this a probable product.	3; genotoxicity alerts present	5; genotoxicity alerts present and repro/developmental tox
Measured data			
Systemic toxicity	-	-	-
Genotoxicity	**Negative** Ames test result, *in vitro* mammalian gene mutation, *in vivo* micronucleus assay	*In vitro* Ames test and mammalian gene mutation **negative**; mammalian chromosome aberration **positive**	-

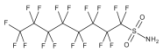
Carcinogenicity		*In vivo* Micronucleus assay **negative**	-
Reproductive & developmental toxicity	-	NOAEL >200 mg/kg bw/day	-
Structural alerts			
Cramer class	Cramer class III	Cramer class III	Cramer class III
Systemic toxicity	-	-	-
Genotoxicity *in vitro*	Non-mutagenicity DNA binding **positive/negative** Ames test **positive/negative**	Non-mutagenicity DNA binding **positive** Ames test **positive/negative**	Non-mutagenicity DNA binding **positive** Ames test **positive/negative**
Genotoxicity *in vivo*	*In vivo* mutagenicity (Micronucleus) alerts by ISS	*In vivo* mutagenicity (Micronucleus) alerts by ISS	*In vivo* mutagenicity (Micronucleus) alerts by ISS
Carcinogenicity	**Positive** **Negative** for (non)genotoxic carcinogenicity	**Negative** for (non)genotoxic carcinogenicity Potential carcinogen	**Negative** for (non)genotoxic carcinogenicity Potential carcinogen
Reproductive toxicity Developmental toxicity	Known precedent reproductive and developmental toxic potential	-	Known precedent reproductive and developmental toxic potential
Category formation			
Rationale: OECD QSAR Toolbox Structural similarity >60% yields 8 analogues, 2 deviate with respect to physico-chemical properties and functional groups. AMBIT reports 3 different analogues with similarity score >0.92 and no measured data. ToxRead finds 6 analogues with a similarity of 0.813 - 0.976, including 95-14-7 (experimental activity: mutagen) and 29385-43-1 (experimental activity: non-mutagen). AIM identified 29385-43-1 as an analogue with measured data. ChemMine finds 10 analogues with similarity >0.6, including 29385-43-1.			
Structural similarity	*Score:*	70-80% / 0.935	80% / 0.976
Functional groups	Benzotriazole Aryl (hetero)arenes Aromatic compound	Benzotriazole Aryl Aromatic compound	Benzotriazole Aryl (hetero)arenes Aromatic compound
Structural alerts		No additional alerts	No additional alerts
Physico-chemical properties	Predicted	Similar	Similar
Lipinski rule	Bioavailable	Bioavailable	Bioavailable
Metabolic pathways		Comparable	Comparable
Read across			
Endpoint:		*Justification:*	
Bacterial gene mutation assay	Positive	1 analogue (95-14-7, self-assessed); all strains and +/-S9 combined, data usage maximum	
Mutagenicity	Negative	6 analogues (assessed by ToxRead) of which 4 are negative.	
QSAR			
Endpoint:			
Mutagenicity	ToxRead and VEGA: negative T.E.S.T. and VEGA: positive	*Predictions are all not considered reliable.*	
Carcinogenicity	VEGA: negative VEGA: positive		
Developmental toxicity	T.E.S.T.: negative VEGA: positive		
Reproductive and developmental toxic potential (DART)	OECD QSAR Toolbox: known precedent VEGA: negative		
Hepatotoxicity	VEGA: positive		

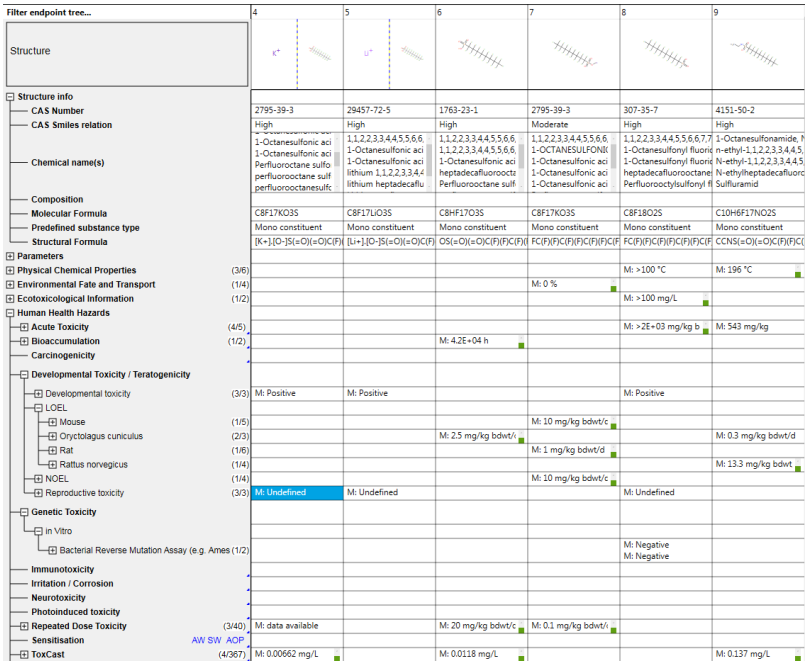
Conclusion: The only experimental data for this substance are negative Ames test results. Few analogues are available. Structural alerts, read across, and QSARs yield equivocal predictions for genotoxicity, carcinogenicity, and reproductive and developmental toxicity. Predicted metabolites contain genotoxicity alerts. Experimental data, in particular

genotoxicity testing results and acceptable daily intake levels based on chronic exposure studies, are required for toxicological evaluation of this substance. Until further information is available, the generic TTC-based drinking water target value of 0.01 µg/L for genotoxic compounds may be applied as a conservative approach.

4.2 New perfluorinated compounds

There is ample information on long-chain perfluorinated chemicals such as PFOS and PFOA with regard to toxicity and environmental behaviour (Post et al., 2017). In recent years, also short-chain perfluorinated compounds have been detected in (sources of) drinking water. The toxicity of these substances has often not been evaluated thoroughly. An example is perfluorooctanesulfonamide (PFOSA), for which no results from chronic toxicity and genotoxicity studies or human health risk assessments were available in the information sources listed in §2.1 and §2.2. In recent BTO research, an analytical method was developed for this substance.

	Target chemical	Analogues
Name CAS SMILES	perfluorooctanesulfonamide 754-91-6 O=S(=O)(N)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	OECD QSAR Toolbox (n=8), AMBIT (n=3), ToxRead (n=3), ChemMine (n=10) and AIM (none with default settings, n=12 with less stringent settings) all provide different analogues.
Metabolites		
Observed (O) / Predicted (P) CYP450 rat S9 rat *in vivo*	No observed metabolites OECD Toolbox and ToxTree: none; Xenosite: not probable OECD Toolbox: none OECD Toolbox: none	
Measured data		
Systemic toxicity Genotoxicity Carcinogenicity Reproductive & developmental toxicity	No measured data for target substance. OECD QSAR Toolbox data for analogues:	



Filter endpoint tree...	4	5	6	7	8	9
Structure						
Structure info						
CAS Number	2795-39-3	29457-72-5	1763-23-1	2795-39-3	307-35-7	4151-50-2
CAS Smiles relation	High	High	High	Moderate	High	High
Chemical name(s)	1-Octanesulfonic aci 1-Octanesulfonic aci Perfluoroctane sulfo perfluorooctane sulf perfluorooctanesulfc	1,1,2,2,3,3,4,4,5,5,6,6, 1-Octanesulfonic aci 1-Octanesulfonic aci lithium 1,1,2,2,3,3,4,4 lithium heptadecaflu	1,1,2,2,3,3,4,4,5,5,6,6, 1,1,2,2,3,3,4,4,5,5,6,6, 1-Octanesulfonic aci heptadecafluoroctz Perfluorooctane sulf	1,1,2,2,3,3,4,4,5,5,6,6, 1-OCTANESULFONIC 1-Octanesulfonic aci 1-Octanesulfonic aci 1-Octanesulfonic aci	1,1,2,2,3,3,4,4,5,5,6,6,7,7 1-Octanesulfonyl fluori 1-Octanesulfonyl fluori heptadecafluorooctane Perfluorooctylsulfonyl fl	1-Octanesulfonamide, ? n-ethyl-1,1,2,2,3,3,4,4,5, N-ethyl-1,1,2,2,3,3,4,4,5, N-ethylheptadecafluoro Sulfluramid
Composition						
Molecular Formula	C8F17KO3S	C8F17LiO3S	C8HF17O3S	C8F17KO3S	C8F18O2S	C10H6F17NO2S
Predefined substance type	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent	Mono constituent
Structural Formula	[K+].[O-]S(=O)(=O)C(F)	[Li+].[O-]S(=O)(=O)C(F)	OS(=O)(=O)C(F)(F)C(F)(	FC(F)(F)C(F)(F)C(F)(F)C(F	FC(F)(F)C(F)(F)C(F)(F)C(F	CCNS(=O)(=O)C(F)(F)C(
Parameters						
Physical Chemical Properties (3/6)					M: >100 °C	M: 196 °C
Environmental Fate and Transport (1/4)				M: 0 %		
Ecotoxicological Information (1/2)					M: >100 mg/L	
Human Health Hazards						
Acute Toxicity (4/5)					M: >2E+03 mg/kg b	M: 543 mg/kg
Bioaccumulation (1/2)			M: 4.2E+04 h			
Carcinogenicity						
Developmental Toxicity / Teratogenicity						
Developmental toxicity (3/3)	M: Positive	M: Positive			M: Positive	
LOEL						
Mouse (1/5)				M: 10 mg/kg bdwt/c		
Oryctolagus cuniculus (2/3)			M: 2.5 mg/kg bdwt/			M: 0.3 mg/kg bdwt/d
Rat (1/6)				M: 1 mg/kg bdwt/d		
Rattus norvegicus (1/4)						M: 13.3 mg/kg bdwt
NOEL (1/4)				M: 10 mg/kg bdwt/c		
Reproductive toxicity (3/3)	M: Undefined	M: Undefined			M: Undefined	
Genetic Toxicity						
in Vitro						
Bacterial Reverse Mutation Assay (e.g. Ames (1/2)					M: Negative M: Negative	
Immunotoxicity						
Irritation / Corrosion						
Neurotoxicity						
Photoinduced toxicity						
Repeated Dose Toxicity (3/40)	M: data available		M: 20 mg/kg bdwt/c	M: 0.1 mg/kg bdwt/		
Sensitisation AW SW AOP						
ToxCast (4/367)	M: 0.00662 mg/L		M: 0.0118 mg/L			M: 0.137 mg/L

Structural alerts	
Cramer class	Cramer class III
Systemic toxicity	-
Genotoxicity *in vitro*	DNA binding **negative**; Ames test **positive/negative**
Genotoxicity *in vivo*	DNA alert for CA and MNT; *In vivo* mutagenicity (Micronucleus) alerts **negative**
Carcinogenicity	No alert for (non)genotoxic carcinogenicity

Reproductive toxicity Developmental toxicity	Not known precedent reproductive and developmental toxic potential	
Category formation		
Rationale: The OECD QSAR Toolbox shows 6 analogues (structural similarity >90%, comparable functional groups and structural alerts) with measured data. Predicted physchem properties are similar, except for analogue #8 and #9 – these are more persistent and have a higher LogKow and lower water solubility, and can thus be included for worst case predictions. ToxRead finds 3 analogues with a similarity of 0.72- 0.83 and negative experimental mutagenicity data.		
Structural similarity	*Score:*	
Functional groups	Alkyl fluoride Alkyl halide Halogen derivative Sulfonamide Sulfonic acid derivative	
Structural alerts	Halogens ; TTC carcinogens category	
Physico-chemical properties	Predicted	
Lipinski rule	Not bioavailable	
Metabolic pathways	-	
Read across		
Endpoint:		*Justification:*
Developmental toxicity	Positive	OECD QSAR Toolbox: 3 analogues with positive test result; 3 other analogues with LOEL in mg/kg range
Bacterial gene mutation assay	Negative	OECD QSAR Toolbox: 1 analogue with negative Ames test data
Mutagenicity	Negative	3 analogues (assessed by ToxRead) which all are negative
QSAR		
Endpoint:		
Mutagenicity	VEGA, T.E.S.T. and ToxRead: negative	*VEGA and ToxRead predictions are not considered reliable.*
Carcinogenicity	VEGA: positive	
Developmental toxicity	VEGA: negative	
Reproductive and developmental toxic potential (DART)	OECD QSAR Toolbox: not known precedent; VEGA: negative	
Hepatotoxicity	VEGA: unknown	

Conclusion: There are no experimental data for this substance and no information on (potential) metabolites. There is no consensus on analogues between different tools. Some structural alerts for genotoxicity are reported, but read across and QSARs do not predict mutagenicity. Carcinogenicity predictions are negative. Predictions for developmental toxicity are equivocal. Experimental data are required for toxicological evaluation of this substance. Since the weight of evidence shows that genotoxicity is unlikely, the generic TTC-based drinking water target value of 0.1 µg/L may be applied until further information is available.

4.3 Disinfection byproducts

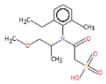
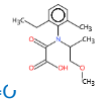
Advanced oxidation processes are important barriers for organic micropollutants in (drinking) water treatment. It is however known that medium pressure UV/H_2O_2 treatment may lead to mutagenicity in the Ames test, which is no longer present after granulated activated carbon (GAC) filtration. Many nitrogen-containing disinfection by-products (N-DBPs) result from the reaction of photolysis products of nitrate with (photolysis products of) natural organic material (NOM) during medium pressure UV treatment of water. The chemical identity and toxicity of most of the N-DBPs are unknown. 3-Nitroindole is one of the N-DBPs that has been identified in recent BTO research.

	Target chemical	Analogues
Name	3-Nitroindole	
CAS	4770-03-0	Very few suitable analogues are identified.
SMILES	C1=CC=C2C(=C1)C(=CN2)[N+](=O)[O-]	
Metabolites		
Observed (O) / Predicted (P)	No observed metabolites	
CYP450	The two most probable metabolites according to OECD Toolbox and Xenosite have no measured data but mutagenicity alerts	
rat S9	The five most probable metabolites according to OECD Toolbox, ToxTree and Xenosite have no measured data but mutagenicity alerts	
rat *in vivo*	The two most probable metabolites according to OECD Toolbox, ToxTree and Xenosite have no measured data but mutagenicity alerts	
Measured data		
Systemic toxicity	-	
Genotoxicity	-	
Carcinogenicity	-	
Reproductive & developmental toxicity	-	
Structural alerts		
Cramer class	Cramer class III	
Systemic toxicity	-	
Genotoxicity *in vitro*	DNA binding, Ames test alerts	
Genotoxicity *in vivo*	*In vivo* mutagenicity (Micronucleus) alerts by ISS	
Carcinogenicity	Positive for genotoxic carcinogenicity	
Reproductive toxicity	Not known precedent reproductive and developmental	
Developmental toxicity	toxic potential	
Category formation		
Rationale: OECD QSAR Toolbox and AMBIT identify no analogues with structural similarity >70% . ToxRead finds 6 analogues with a similarity score of >0.9 with experimental mutagenic activity. AIM identifies one analogue, which is not reported by ToxRead.		
Structural similarity	*Score:*	
Functional groups	Anion Aromatic compound Cation Heterocyclic compound Nitro compound	
Structural alerts		
Physico-chemical properties	Predicted	
Lipinski rule	Bioavailable	
Metabolic pathways		
Read across		
Endpoint:		*Justification:*
Mutagenicity	Positive	6 mutagenic analogues (assessed by ToxRead)
QSAR		
Endpoint:		
Mutagenicity	ToxRead, T.E.S.T. and VEGA: positive	*ToxRead, T.E.S.T. and VEGA predictions are considered reliable.*
Carcinogenicity	VEGA: negative (not reliable) VEGA: positive (reliable)	
Developmental toxicity	VEGA: positive (not reliable)	
Reproductive and developmental toxic potential (DART)	OECD QSAR Toolbox: not known precedent VEGA: negative (not reliable)	
Hepatotoxicity	-	

Conclusion: The CAS number was not recognised by any of the tools and the substance is not present in the underlying databases. Very few analogues are identified. Structural alerts, read across, and QSARs all point to mutagenic activity and potential for genotoxic carcinogenicity. The most probable metabolites show structural alerts for genotoxicity as well. Genotoxic potential needs to be confirmed by experimental testing. In addition, an acceptable daily intake level or virtually safe dose needs to be determined. Until further information is available, the generic TTC-based drinking water target value of 0.01 µg/L for genotoxic compounds may be applied.

4.4 Metabolites of plant protection products

In European and Dutch drinking water standards, metabolites of plant protection products are divided in metabolites that are relevant or metabolites that are non-relevant for human health. This classification is based on structural characteristics, biological activity and (geno)toxicity. Since full human health risk assessment of metabolites is not a standard part of plant protection product regulation (and not performed for minor metabolites), information on potential metabolites and toxicological data is often lacking or incomplete. An example is metolachlor, two metabolites of which are often detected in (sources of) drinking water.

	Target chemical	**Target chemical**
Name **CAS** **SMILES**	Metolachlor-ESA 171118-09-5 CCc1cccc(C)c1N (C(C)COC)C(=O)CS(O)(=O)=O 	Metolachlor-OA 152019-73-3 CCc1cccc(C)c1N (C(C)COC)C(=O)C(O)=O 
Metabolites		
Observed (O) / Predicted (P)	No observed metabolites	No observed metabolites
Measured data		
Systemic toxicity	-	-
Genotoxicity	-	-
Carcinogenicity	-	-
Reproductive & developmental toxicity	-	-
Structural alerts		
Cramer class	Cramer class III	Cramer class III
Systemic toxicity	-	-
Genotoxicity *in vitro*	Non-mutagenicity DNA binding **negative** Ames test **negative**	Non-mutagenicity DNA binding **negative** Ames test **negative**
Genotoxicity *in vivo*	*In vivo* mutagenicity (Micronucleus) alerts by ISS	*In vivo* mutagenicity (Micronucleus) alerts by ISS
Carcinogenicity	**Negative** for (non)genotoxic carcinogenicity	**Negative** for (non)genotoxic carcinogenicity
Reproductive toxicity Developmental toxicity	Known precedent reproductive and developmental toxic potential	Known precedent reproductive and developmental toxic potential
Category formation		
Rationale: OECD QSAR Toolbox identifies no analogues with structural similarity >70%. The 10 most suitable analogues according to ToxRead have a similarity of 0.78 - 0.85; 7 of them are experimental non-mutagens. ChemMine yields different analogues which all have a lower similarity score. AIM identifies one other analogue with less stringent settings. *Rationale:* OECD QSAR Toolbox identifies no analogues with structural similarity >70%. The 10 most suitable analogues according to ToxRead have a similarity of 0.7 - 0.9; 6 of them are experimental non-mutagens. ChemMine yields different analogues which all have a lower similarity score. AIM finds no analogues.		
Structural similarity	*Score:*	
Functional groups	Alkane, branched with secondary carbon	Alkane, branched with

	Alkyl (hetero)arenes Aromatic compound Carboxylic acid tert. amide Dialkylether Sulfonic acid derivative	secondary carbon Alkyl (hetero)arenes Aromatic compound Organic amide and thioamide Dialkylether
Structural alerts	Small alkyl toluene derivatives	Small alkyl toluene derivatives
Physico-chemical properties	Predicted	Predicted
Lipinski rule	Bioavailable	Bioavailable
Metabolic pathways		
Read across		
Endpoint:		*Justification:*
Mutagenicity	Negative	Predicted for both substances by ToxRead based on 10 analogues.
QSAR		
Endpoint:		
Mutagenicity	ToxRead, ToxTree, T.E.S.T. and VEGA: negative (not reliable)	ToxRead, ToxTree, T.E.S.T. and VEGA: negative (not reliable)
Carcinogenicity	ToxTree and VEGA: positive (not reliable) VEGA: negative (not reliable)	VEGA: positive (not reliable) ToxTree and VEGA: negative (not reliable)
Developmental toxicity	T.E.S.T. and VEGA: positive (not reliable)	VEGA: negative (moderate reliability) T.E.S.T.: positive (not reliable)
Reproductive and developmental toxic potential (DART)	OECD QSAR Toolbox: known precedent VEGA: negative (not reliable) VEGA: positive (not reliable)	OECD QSAR Toolbox: known precedent VEGA: negative (not reliable)
Hepatotoxicity	VEGA: negative	VEGA: negative (moderate reliability)

Conclusion: There are no experimental data for these metabolites. None or only moderately similar analogues are identified by the applied tools; read across predictions are therefore not reliable. QSARs do not provide reliable toxicity predictions as well. Experimental data are thus required for toxicological evaluation of these substances. Until further information is available, the generic TTC-based drinking water target value of 0.01 µg/L may be applied as a conservative approach.

5 Conclusions and final remarks

In toxicological evaluations of chemical drinking water contaminants, some types of data prevail over other. Validated measured *in vivo* toxicity data from a well-designed laboratory study are always preferred. When no experimental toxicity data are available, data on appropriate analogues may be used to perform read across. If no toxicity data on the chemical or an appropriate analogue are available, data may be predicted by appropriately using scientifically sound (*in silico*) models.

This report presents *in silico* tools that can aid in structural profiling, read across based on analogues, and QSAR analysis. This is not an exhaustive overview: we only focused on freely available tools that yield information on human health hazards, with emphasis on genetic, reproductive and developmental, and chronic toxicity (i.e. not a full toxicological profile and no environmental behavior of chemicals). More tools are available and will certainly be developed in future; *in silico* methods are likely to expand to include models for specific types of toxicity and chemicals, provide insight into toxicological pathways, and combine and compare results from different models. In addition, models will be customized to meet users' demands and refined when new data become available (Raies and Bajic, 2016). **Continuous inventarisation** is thus warranted. The OECD QSAR Toolbox currently includes one of the largest collections of publicly available data and the most extensive range of analysis options of the currently available tools. In addition, training in appropriate use of this software is provided by the developers.

Expertise is needed to perform and evaluate predictions derived from *in silico* tools: the user is ultimately responsible for the assessment. Even though state of the art tools are used, the validity of the (Q)SAR models and underlying databases, applicability domain, prediction of physicochemical properties and kinetics, similarity of chemicals, cut-off points etc. always need to be critically evaluated. Especially selection of suitable analogues with experimental data for read across approaches requires careful consideration of all characteristics of the chemicals. This is time-consuming, and therefore not feasible for large sets of chemicals. In addition, a prediction approach that proves appropriate for a certain chemical and/or toxicological endpoint may not yield (proper) results for other substances or effects. When using *in silico* tools, it is advised to take care of transparent **documentation** of the applied procedure, data used, uncertainty analysis, and decisions made. Appendix IV can assist in this.

It should be noted that toxicity predictions generally concern **hazard identification** and not risk characterization (acceptable daily intake levels are for instance not calculated) and cannot replace experimental toxicity testing. QSAR prediction models are currently not considered reliable for complex toxicological endpoints. Negative predictions in particular need to be substantiated by additional information, especially when they are based on a limited set of structural alerts. The **adverse outcome pathway** (AOP) concept is foreseen to substantiate mechanistic plausibility of toxicity predictions by enabling comparison of analogues with respect to key molecular events causally linked to the toxicological endpoint of interest. There are currently a limited number of endpoints for which AOPs have been formally developed, but these will become more and more available (Schultz et al., 2015). **Guidance documents** on toxicological evaluation using *in silico* tools are available and expected to be published by health protection authorities in the future.

The main advantages of all *in silico* tools are i) quick **screening** of chemical hazard (which may trigger further testing) and ii) providing supporting information of different nature and from different sources. *In silico* models are complementary themselves as well (Baken and Kools 2014). As the examples in chapter 4 show, different tools may generate different results. **Multiple** (non)testing approaches, tools, and models should thus be used in parallel. Agreement among predictions generated by independent and scientifically valid tools increases the confidence in the predictions made. A **weight of evidence approach,** in which needs all available (non-)testing information is gathered and compared, needs to be applied for an overall assessment of the support of toxicity predictions. Currently, relatively little experience with this type of data integration is available, and no formal guidance has as yet been provided for this. Nevertheless, *is silico* tools are regarded as a useful component of the toxicity assessment process (Raies and Bajic, 2016).

References

Autiero, M. (ToxMinds BVBA) Overview of Publicly Available In Silico Tools in Supporting Hazard Assessments. CAAT Academy Hands-on training 'Use of In Silico Tools in Chemical's Hazard Assessment', Brussels, 5-6 October 2017.

Baken, K.A. and Kools, S. (2014) BTO 2014.009 Innovative testing strategies and their relevance for evaluating chemical drinking water quality. KWR Watercycle Research Institute, Nieuwegein.

Baken, K.A., Schriks, M., Sjerps, R.M.A. (2015). BTO 2015.056 Toxicologische risicobeoordeling geprioriteerde stoffen. KWR Watercycle Research Institute, Nieuwegein, The Netherlands.

Baken, K.A. and Sjerps, R.M.A. (2016) BTO 2016.069 The Threshold of Toxicological Concern (TTC) - refinement of the concept and application to drinking water. KWR Watercycle Research Institute, Nieuwegein, The Netherlands.

Baken, K.A. and Dingemans, M.M.L. (2017) BTO 2017.008 Effect-based monitoring with bioassays – a roadmap. KWR Watercycle Research Institute, Nieuwegein.

Baken, K.A., Sjerps, R.M.A., Schriks, M., Van Wezel, A.P. (2018) Toxicological relevance and Threshold of Toxicological Concern (TTC) for drinking water relevant contaminants of emerging concern. *Submitted for publication.*

Boobis A, Brown P, Cronin MTD, Edwards J, Galli CL, Goodman J, Jacobs A, Kirkland D, Luijten M, Marsaux C, Martin M, Yang C, Hollnagel HM. Origin of the TTC values for compounds that are genotoxic and/or carcinogenic and an approach for their re-evaluation. Crit Rev Toxicol. 2017 Sep;47(8):705-727.

Bower, D., Cross, K. P., Escher, S., Myatt, G. J., Quigley, D. P. CHAPTER 9 In silico Toxicology: An Overview of Toxicity Databases, Prediction Methodologies, and Expert Review. In Computational Systems Pharmacology and Toxicology, The Royal Society of Chemistry: 2017; pp 209-242.

ECHA (2008) Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. European Chemicals Agency, Helsinki, Finland.

ECHA (2016) Practical Guide – How to use and report (Q)SARs. Version 3.1. European Chemicals Agency, Helsinki, Finland.

ECHA (2017a) Read-Across Assessment Framework (RAAF). European Chemicals Agency, Helsinki, Finland.

ECHA (2017b) Non-animal approaches. Current status of regulatory applicability under the REACH, CLP and Biocidal Products regulations. European Chemicals Agency, Helsinki, Finland.

EFSA (2012) Scientific Opinion on exploring options for providing advice about possible human health risks based on the concept of Threshold of Toxicological Concern (TTC). EFSA Journal 10(7):2750.

EFSA/WHO (2016). Review of the Threshold of Toxicological Concern (TTC) approach and development of new TTC decision tree. European Food Safety Authority and World Health Organization. EFSA Supporting publication 2016: EN-1006.

Jeliaskova, N. (Ideaconsult Ltd) AMBIT2 Hands-on Training Workshop, Brussels, 29 September 2017.

Kirchmair, J. (University of Hamburg) Predicting Chemical's ADME Behavior. CAAT Academy Hands-on training 'Use of In Silico Tools in Chemical's Hazard Assessment', Brussels, 5-6 October 2017.

Kolkman, A., Vughs, D., Bäuerlein, P, Baken, K., Schriks, M. (2013) BTO 2013.059 Groepsgewijze analyse en beoordeling van stoffen: implementatie van 'structural alerts' in waterkwaliteitsmonitoring. KWR Watercycle Research Institute, Nieuwegein.

Kroes, R., Renwick, A.G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., Schilter, B., Schlatter, J., van Schothorst, F., Vos, J.G., Würtzen, G. (2004) Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem Toxicol 42(1):65-83.

Mons, M.N., Heringa, M.B., van Genderen, J., Puijker, L.M., Brand, W., van Leeuwen, C.J., Stoks, P., van der Hoek, J.P., van der Kooij, D. (2013) Use of the Threshold of Toxicological Concern (TTC) approach for deriving target values for drinking water contaminants. Water Research 47, 1666-1678.

Munro, I.C., Ford, R.A., Kennepohl, E., Sprenger, J.G. (1996) Correlation of structural class with no-observed-effect levels: A proposal for establishing a threshold of concern. Food and Chemical Toxicology 34, 829-867.

Petry, T. (ToxMinds BVBA) Process to In Silico - Based Chemical Hazard Assessment. CAAT Academy Hands-on training 'Use of In Silico Tools in Chemical's Hazard Assessment', Brussels, 5-6 October 2017.

Post GB, Gleason JA, Cooper KR. (2017) Key scientific issues in developing drinking water guidelines for perfluoroalkyl acids: Contaminants of emerging concern. PLoS Biol. 20;15(12):e2002855.

Raies, A. B., & Bajic, V. B. (2016). In silico toxicology: computational methods for the prediction of chemical toxicity. Wiley Interdisciplinary Reviews. Computational Molecular Science, 6(2), 147–172.

Rulis, A. (1986) De minimis and threshold of regulation.In: Food Protection Technology. C.W. Felix, ed. Lewis Publishers, pp 29-37.

SCCS/SCHER/SCENIHR (2012) Joint opinion on use of the Threshold of Toxicological Concern (TTC) approach for human safety assessment of chemical substances with focus on cosmetics and consumer products. SCCP/1171/08.

Schriks M, Heringa MB, van der Kooi MM, de Voogt P, van Wezel AP (2010). Toxicological relevance of emerging contaminants for drinking water quality. Water Res. 2010 Jan;44(2):461-76.

Schultz, T. W., Amcoff, P., Berggren, E., Gautier, F., Klaric, M., Knight, D. J., Mahony, C.; Schwarz, M., White, A., Cronin, M. T. D. (2015) A strategy for structuring and reporting a read-across prediction of toxicity. Regulatory Toxicology and Pharmacology 72, (3), 586-601.

Attachment I

CAAT academy hands-on training

CAAT academy

Use of In Silico Tools in Chemical's Hazard Assessment

5th - 6th October 2017

PROGRAM

Thursday, 5th October 2017

08:30 - 08:50		Arrival, Registration, Coffee
08:50 - 09:00	Introduction	CAAT Academy Overview *François Busquet (CAAT Academy)*
09:00 - 09:30	Lecture	Process to In Silico - Based Chemical Hazard Assessment *Thomas Petry (ToxMinds BVBA)*
09:30 - 10:15	Lecture	Overview of Publicly Available In Silico Tools in Supporting Hazard Assessments *Monica Autiero (ToxMinds BVBA)*
10:15 - 11:00	Lecture	Use of In Silico Tools in a Regulatory Context *Katrin Schütte (European Commission)*
11:00 - 11:20		Coffee Break
11:20 - 11:45	Lecture	Use of the Ambit Tool in Identifying Analogues for Read - Across *Nina Jeliazkova (Ideaconsult Ltd)*
11:45 - 12:15	Lecture	Predicting Chemical's ADME Behavior *Johannes Kirchmair (University of Hamburg)*
12:15 - 12:45	Discussion	Questions and Answers *All*
12:45 - 13:45		Lunch
13:45 - 15:45	Hazard & Analogue ID Tools Training	OECD Toolbox, AIM, ToxTree; ChemMine Vega Ambit *All Presenters*
15:45 - 16:00		Coffee Break
16:00 - 17:30	ADME Tools Training	OECD Toolbox, MetaPrint 2D, SMARTCyp *All Presenters*
19:00		Social Dinner

Friday, 6th October 2017

HOT: Hands-on Training

08:00 - 08:30		Arrival, Coffee
08:30 - 10:30	HOT Case Study I	Toxicological Endpoint Data Gap Case *All Participants and Presenters*
10:30 - 10:45		Coffee Break
10:45 - 12:45	HOT Case Study II	Analogue - Based Read Across (Focus: ADME) *All Participants and Presenters*
12:45 - 13:45		Lunch
13:45 - 15:45	HOT Case Study III	Use of In Silico Tools to Support Mode of Action (MoA) Based Hazard Assessment *All Participants and Presenters*
15:45 - 17:00	Discussion	Presentation of Case Study Results and Discussion *All Participants and Presenters*
17:00 - 17:15		Closure of the Training - Hanging out of the Certificates

toxminds Product Safety & Regulatory Affairs

IDEA CONSULT thinking ahead

UHH

www.caat-academy.org

Attachment II

CEFIC-LRI hands-on Training

presentation materials are published at http://cefic-lri.org/toolbox/ambit/

cefic

LRI Long-Range Research Initiative

AGENDA

AMBIT2 Hands-on Training Workshop

"Cefic LRI AMBIT2 with IUCLID6 support and extended search capabilities"

Venue **CEFIC Offices in Brussels (Av Van Nieuwenhuyse 4, 1160 Brussels)**
September 29th 2017, 10.30 – 16.00 h

Topics

1. **10.30 – 10.40 Welcome to the CEFIC LRI AMBIT2 training Workshop** (Bruno Hubesch)

2. **10.40 – 10.55 AMBIT is good for you!** (Qiang Li, Clariant)

3. **10.55 – 11.15 AMBIT2 Project Overview** (Nina Jeliaskova, Ideaconsult Ltd)
   - Accessibility of IUCLID6 substance data
   - Extracting relevant substance data from IUCLID6 using data filters
   - Use of the LRI Chemoinformatic System AMBIT2 to assign structures to substance constituents, impurities and additives
   - Implement workflows for Read across of Substance endpoint data and Category formatior

4. **11.15 – 11.30 IUCLID Substance Data** (Nikolay Kochev, Ideaconsult Ltd)
   - IUCLID6 Substance Identity Concept: Characterisation of a substance using constituents, impurities and additives
   - Extracting data from IUCLID6 using filters which can be fine tuned

5. **11.30 – 12.05 AMBIT2 Chemoinformatic System** (Nina Jeliazkova/ Nikolay Kochev, Ideaconsult Ltd)
   - Enhancing AMBIT2 to allow import of IUCLID6 substance data
   - Implementation of different search functionalities related to structures, substances an endpoint data
   - AMBIT2 user management system to grant access rights via roles
   - IT Technique necessary to run AMBIT2

6. **12.05 – 12.25 Assessment Workflows for Read across and Substance Category formation** (Qiang Li, Clariant)
   - General aspects on non-testing approaches
   - How to support an assessor in establishing a justification for a read across / category formation
   - Elements of the read across / category formation workflow
   - Steps in the workflow
   - Reporting

7. **12.10 – 13.10 Lunch**

8. **13.10 – 14.00 AMBIT2 Demonstration** (Qiang Li, Clariant)

9. **14.00 – 14.20 Coffee Break + Networking**

10. **14.20 – 16.00 Hands on AMBIT functionality using own notebook (all interested)**

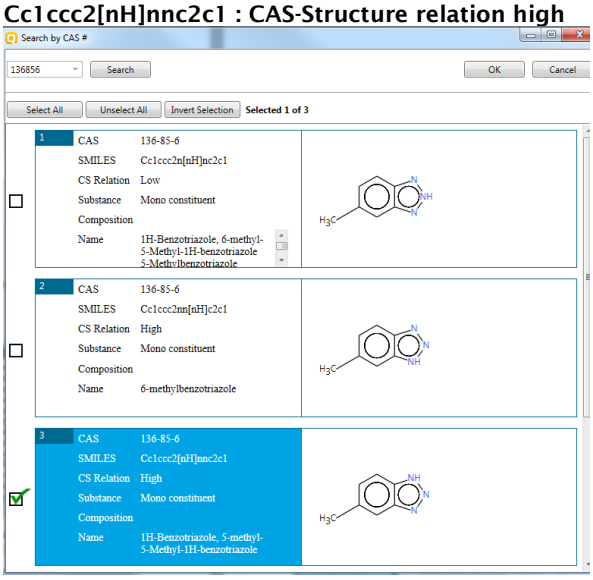
Attachment III

Application of *in silico* tools for an example chemical

Name: 1H-Benzotriazole, 5-methyl-
CAS: 136-85-6

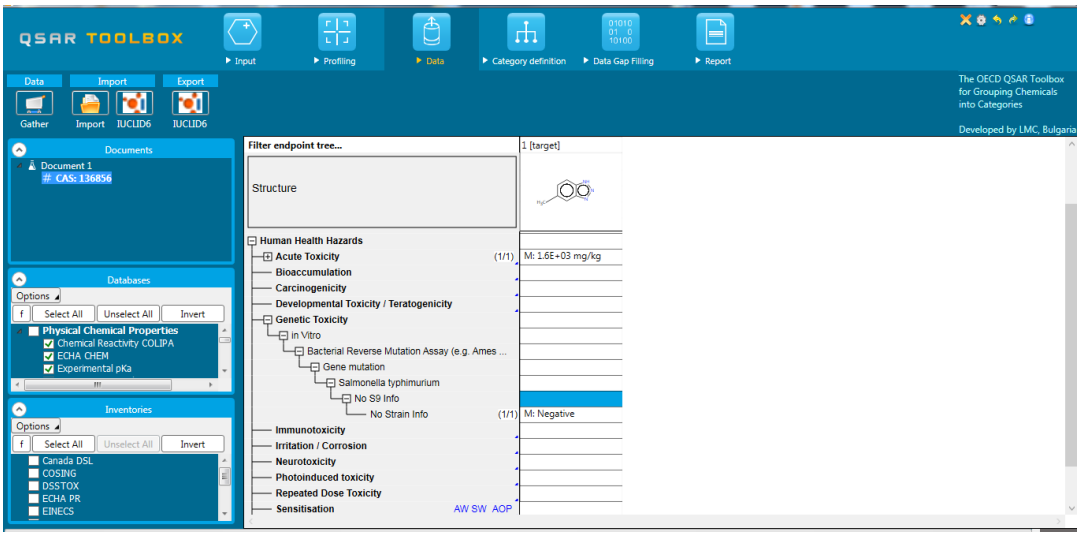
OECD QSAR Toolbox

Identity
OECD Toolbox shows 3 chemical structures based on CAS:
Cc1ccc2n[nH]nc2c1: CAS-Structure relation low
Cc1ccc2nn[nH]c2c1 : CAS-Structure relation high
Cc1ccc2[nH]nnc2c1 : CAS-Structure relation high



PubChem: CAS linked to 2D = CC1=CC2=NNN=C2C=C1
ChemID plus: CAS linked to c12c(cc(C)cc2)nn[nH]1
Chemistry Dashboard: CAS linked to CC1=CC2=C(NN=N2)C=C1

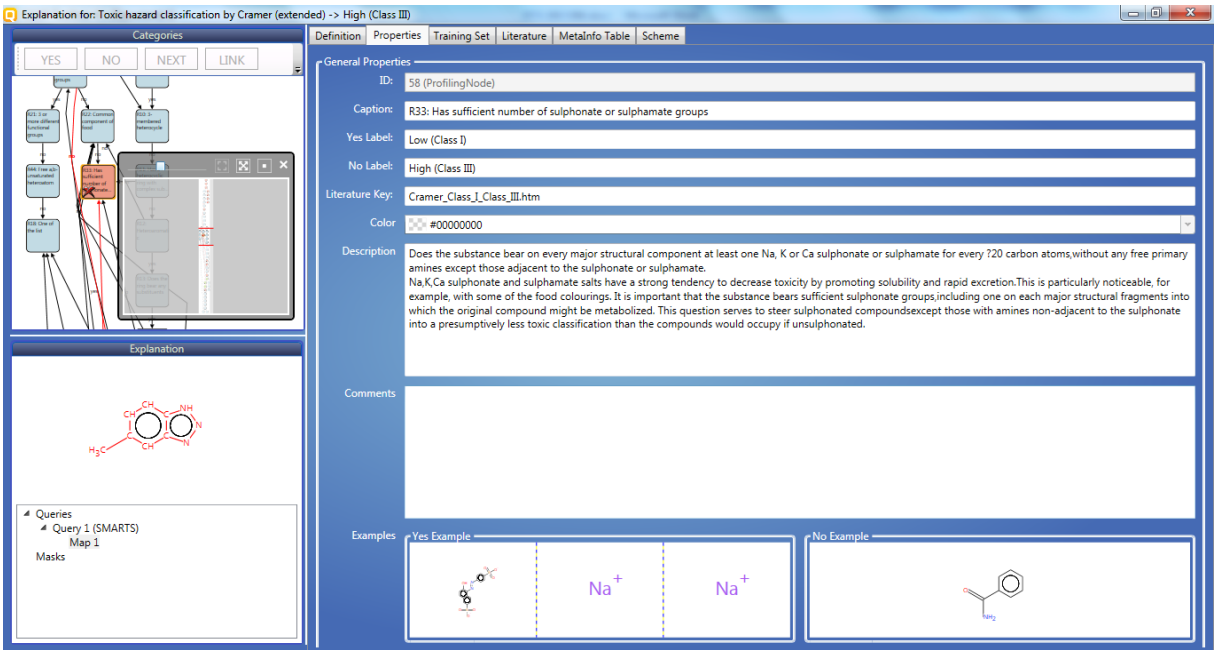
Measured data:
One negative Ames test study result is reported (strain and metabolism not specified).



Profiling

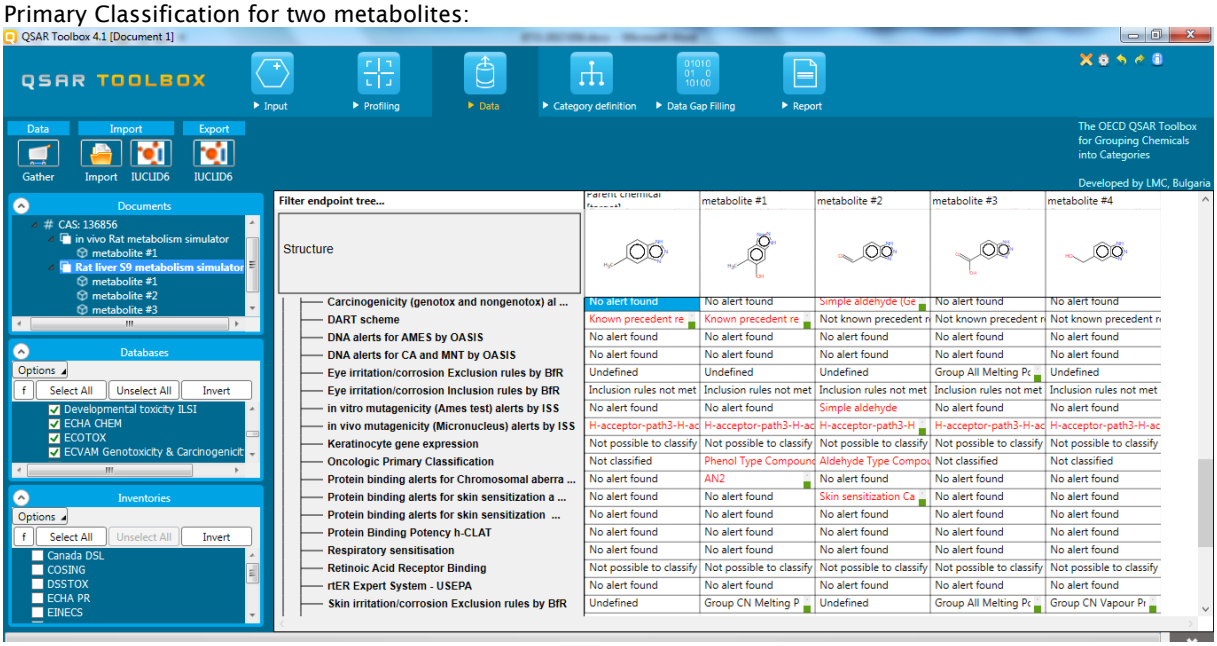
- Cramer class III
- *in vivo* mutagenicity (Micronucleus) alerts by ISS:

- Known precedent reproductive and developmental toxic potential / Toluene and small alkyl toluene derivatives (8a):



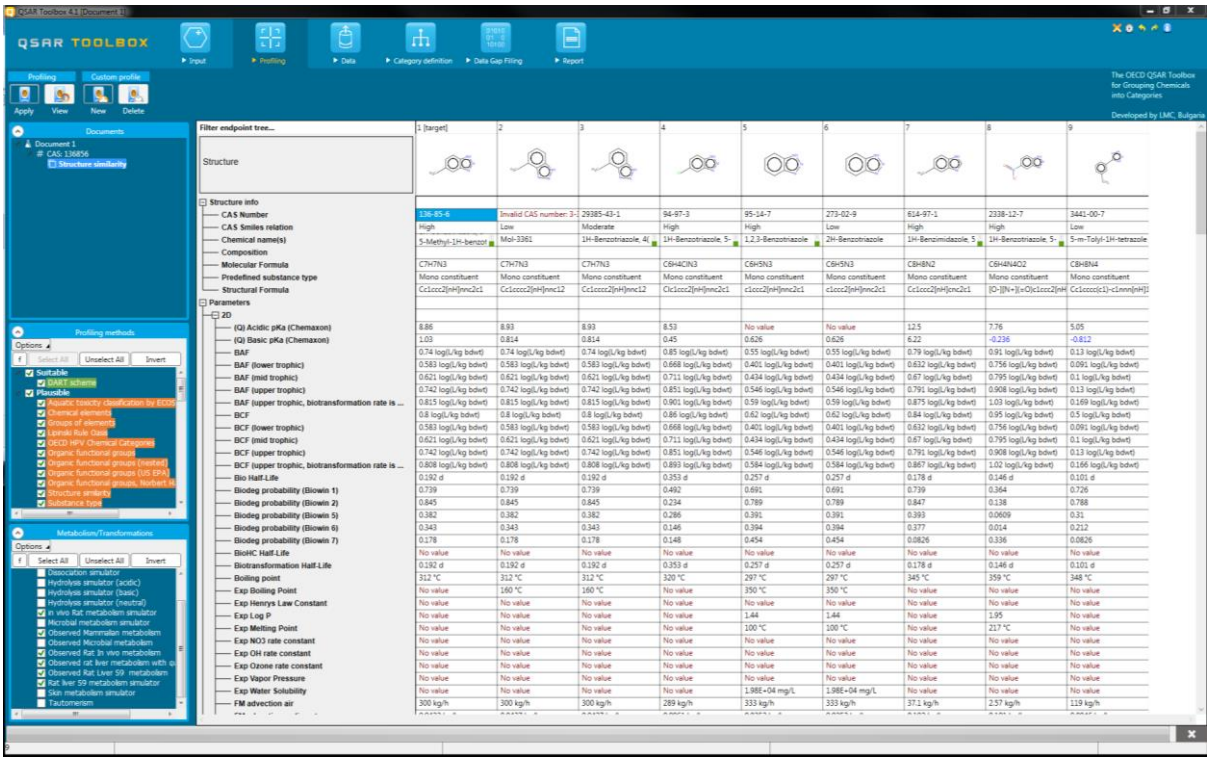
Metabolites

No observed metabolites are reported. Simulated rat S9 metabolites yields 4 metabolites; the first one is also predicted for rat *in vivo* metabolism. No measured data are available for the predicted metabolites. All show the same *in vivo* mutagenicity (Micronucleus) alerts by ISS as the parent. In addition, DNA binding by OECD (Michael addition >> P450 Mediated Activation to Quinones and Quinone-type Chemicals >> Alkyl phenols) is predicted for one metabolite, *in vitro* mutagenicity (Ames test) alerts by ISS for one other metabolite, and Oncologic Primary Classification for two metabolites:



Category definition

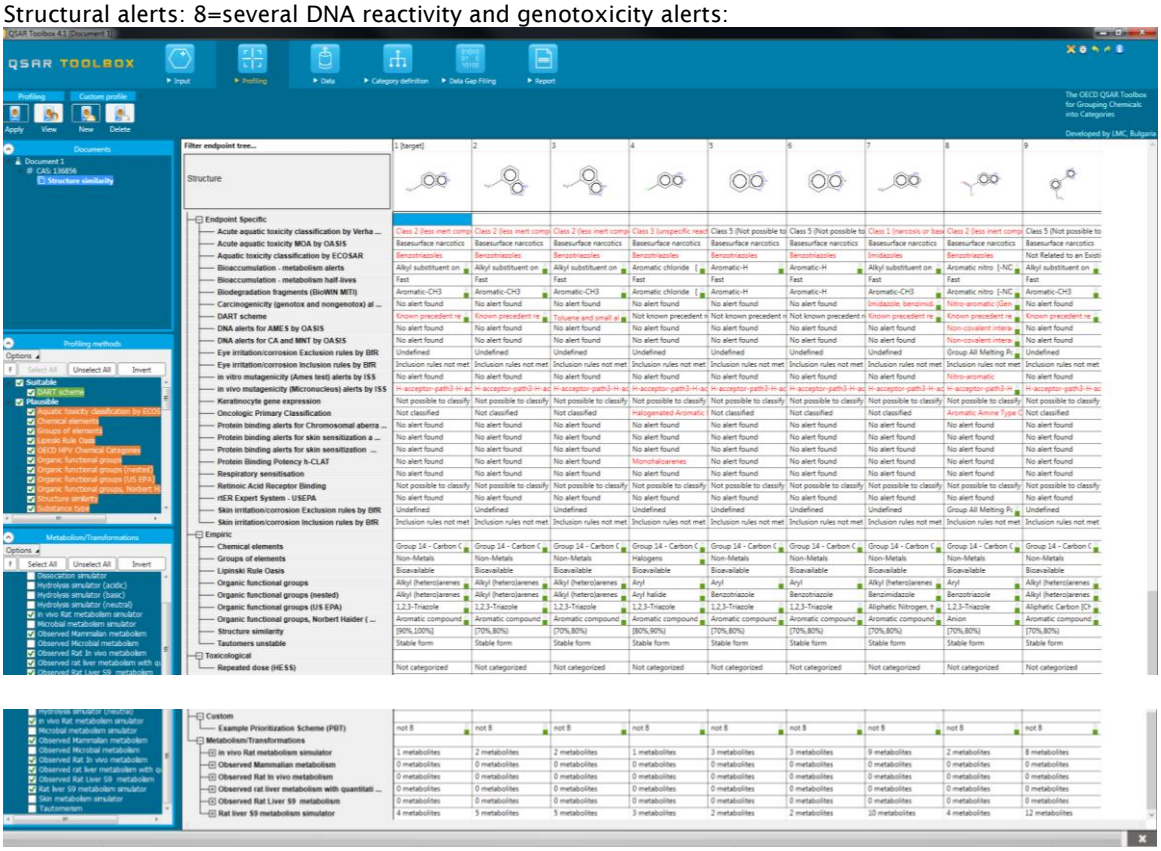
Structural similarity >60% yields 8 analogues:



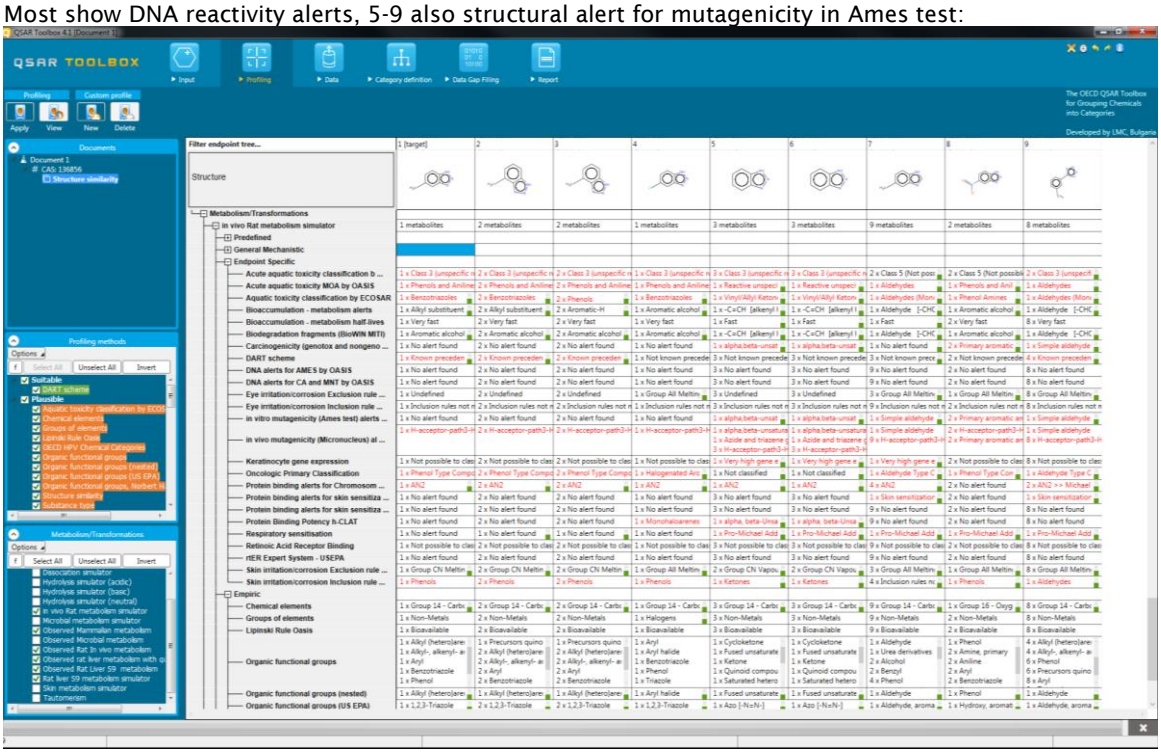
Physico-chemical properties: mainly predicted (instead of measured); all are bioavailable; 4-8-9 deviate.

Functional groups : all Cramer class III; benzotriazoles, 4=Halogen (CL group), 7 = imidazole, 8=nitro-aromatic amine, 9 = not categorized.

Structural alerts: 8=several DNA reactivity and genotoxicity alerts:

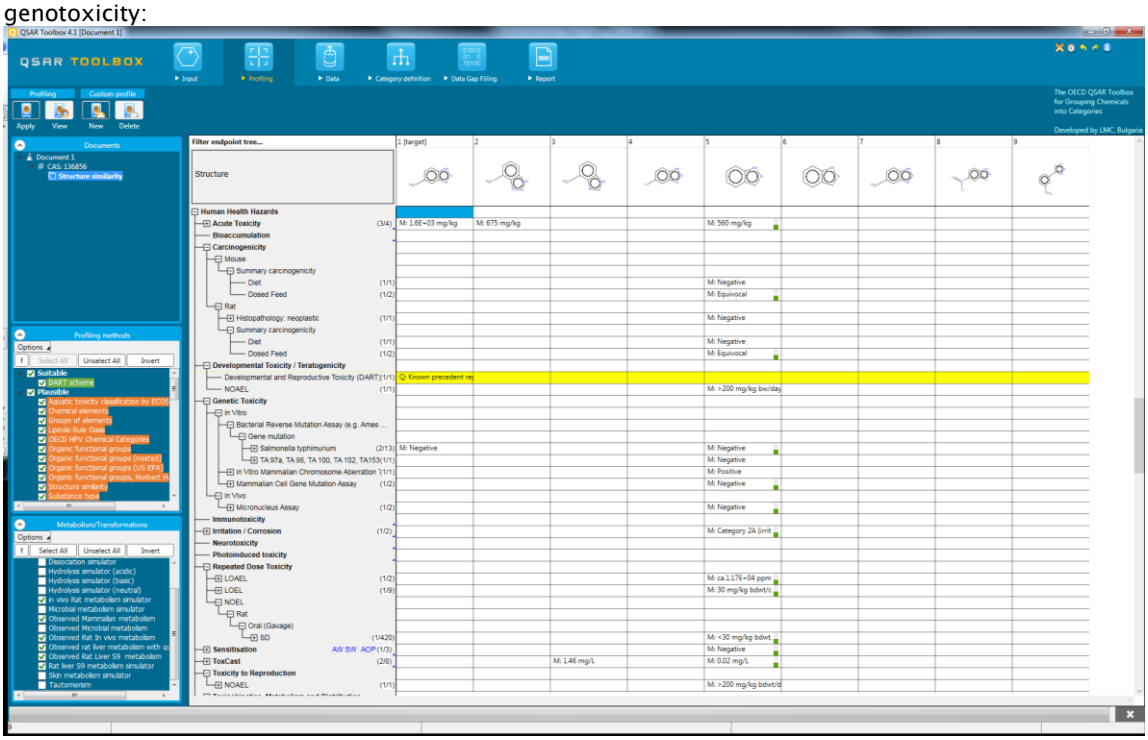


Metabolites: no observed metabolites; rat in vivo simulator: several metabolites for 7 and 9.
Most show DNA reactivity alerts, 5-9 also structural alert for mutagenicity in Ames test:

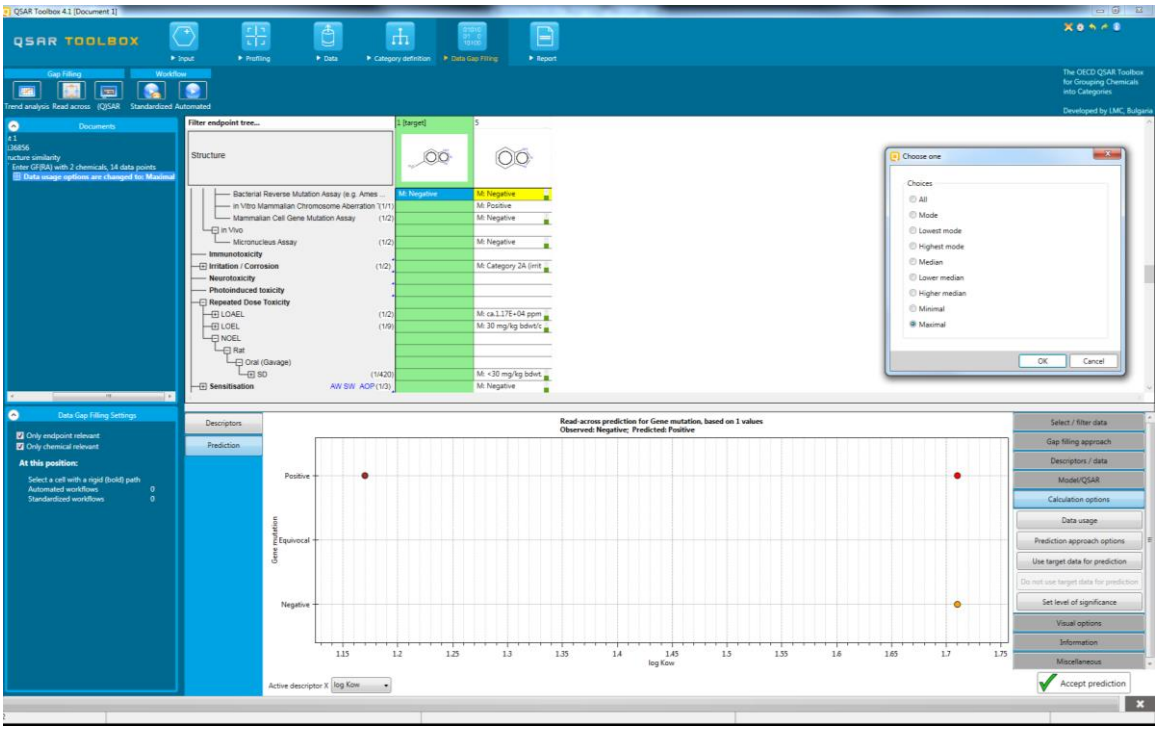


Read across

Measured data: only analogue 5 has measured data for developmental toxicity and genotoxicity:

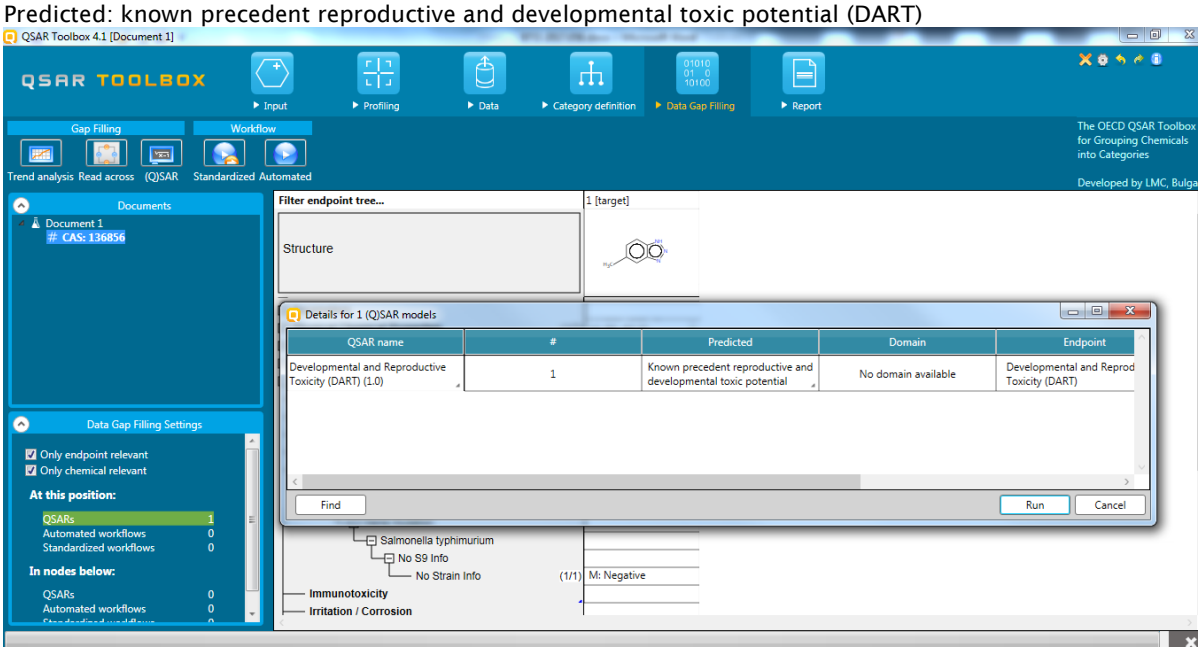


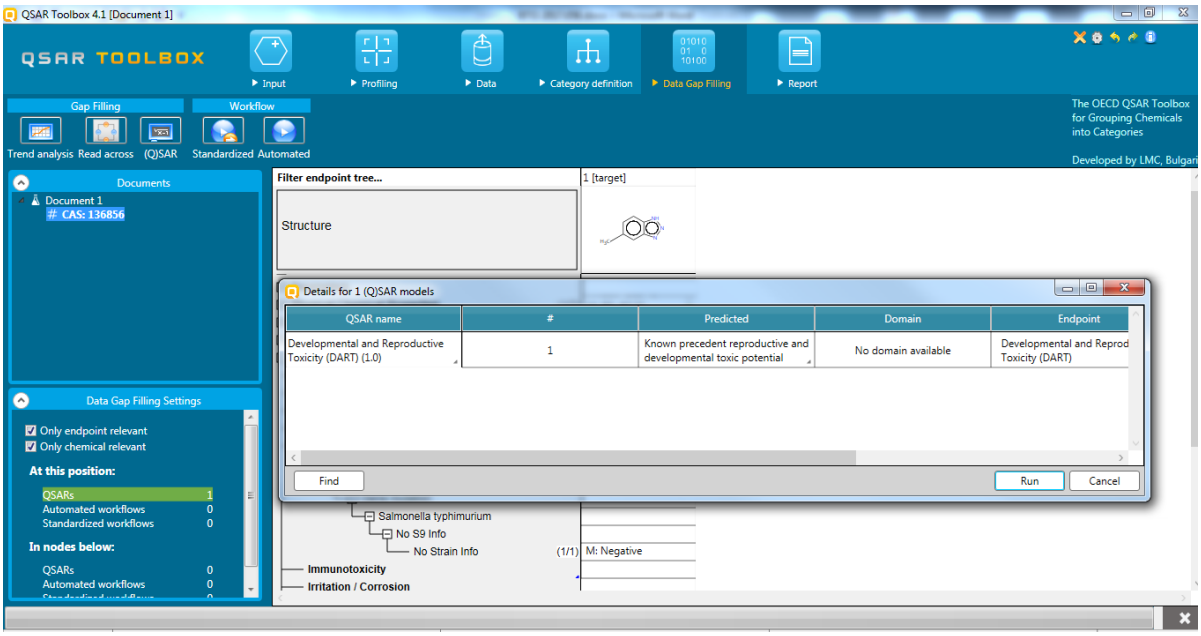
Data gap filling for Bacterial gene mutation assay, all strains and +/-S9 combined, data usage maximum: prediction = positive:



QSAR

Predicted: known precedent reproductive and developmental toxic potential (DART)



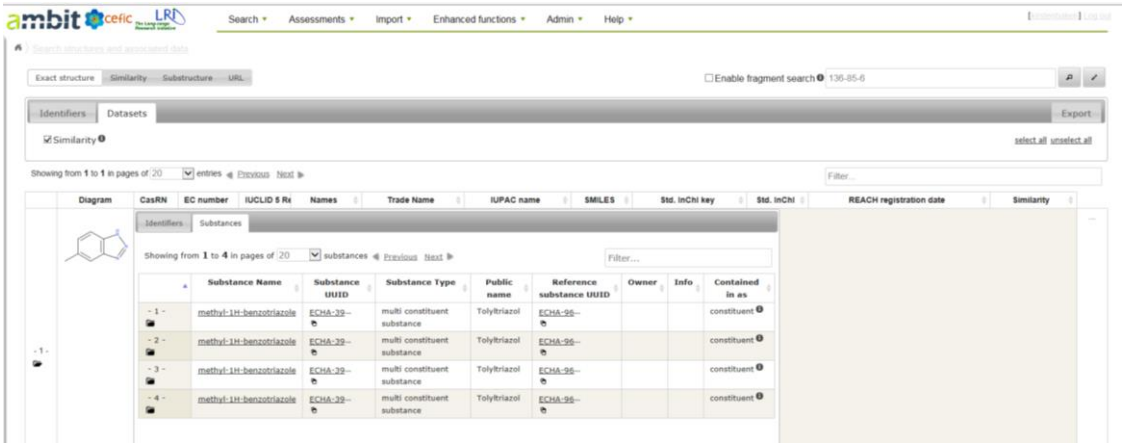


AMBIT

<u>Identity</u>

N1=NC=2C=C(C=CC2N1)C

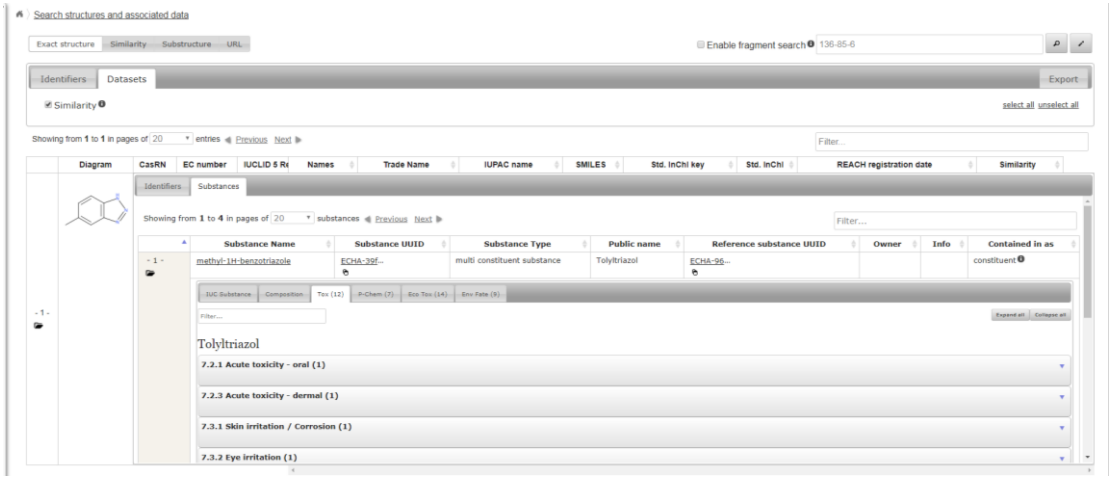
Constituent of 4 ECHA substances:



<u>Measured data</u>

All 4 substances: subacute NOAEL 150 mg/kg bw/day (nominal), reproduction generation P and developmental NOAEL >200 mg/kg bw/day.

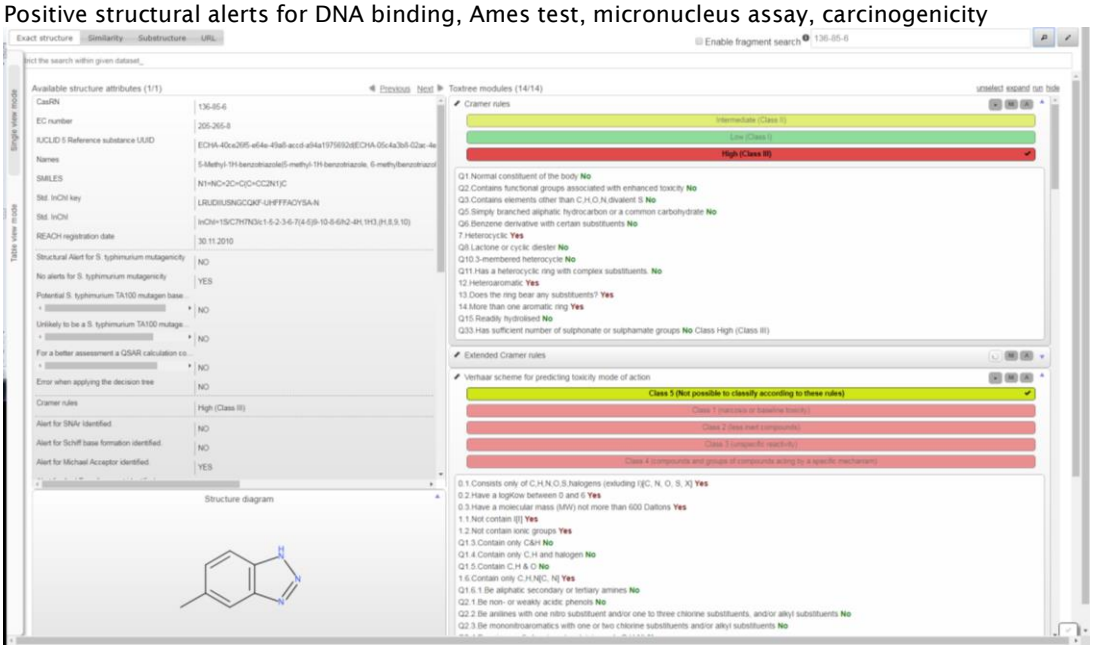
Negative for bacterial reverse mutation assay, *in vitro* mammalian gene mutation, *in vivo* micronucleus assay:



Profiling

Cramer class III

Positive structural alerts for DNA binding, Ames test, micronucleus assay, carcinogenicity



Metabolites

No metabolites are reported in related substances.

Category definition

Three analogues are found; one is a constituent in a multi constituent substance, for the two other no substances are reported.

Home

Assessment identifier | Collect structures | Endpoint data used | Assessment details | Report

Collect structures | List collected

Search

Exact structure | Similarity | Substructure | URL

Enable fragment search Cc1ccc2[nH]nnc2c1

Showing from 1 to 4 in pages of 20 entries Previous Next

Filter...

Diagram	CasRN	EC number	Names	Similarity	Rationale
	136-85-6	205-265-8	6-methylbenzotriazole | 1H-Benzotriazole, 5-methyl- | , 5-Methyl-1H-benzotriazole | 5-methyl-1H-benzotriazole	1	Reason for selection_
	41253-36-5	255-281-4	sodium 5-methyl-1H-benzotriazolide | 1H-Benzotriazole, 5-methyl-, sodium salt | , sodium 5-methylbenzotriazol-1-ide	1	Reason for selection_
	4184-79-6	224-058-3	5,6-dimethyl-1H-benzotriazole	0.97	Reason for selection_
	67924-12-3	267-797-7	bis(5-methyl-1H-benzotriazole) sulphate	0.92	Reason for selection_

Some measured physicochemical properties are reported for the related substances:

Home

Assessment identifier | Collect structures | Endpoint data used | Assessment details | Report

Diagram	CasRN	EC number	Names	Similarity	Rationale

Identifiers | Substances

Name	Endpoint	Result	Text	Guideline	Owner	Citation	Reliability	UUID
Disseminated endpoint study record	log Pow	1.714	-			true	2 (reliable with restrictions)	ECHA-93...
Disseminated endpoint study record	log Pow	(1.079, 1.083]	-	OECD Guideline 117 (Partition Coefficient (n-octanol / water), HPLC Method)		true	1 (reliable without restriction)	ECHA-b9...

Showing 2 substance(s) (1 to 2) Previous Next

4.8 Water solubility (1)

Identifiers | Substances

Name	Endpoint	Result	Text	Guideline	Owner	Citation	Reliability	UUID
Disseminated endpoint study record	log Pow	(1.083, 1.091]	-	OECD Guideline 117 (Partition Coefficient (n-octanol / water), HPLC Method)		true	1 (reliable without restriction)	ECHA-22f...

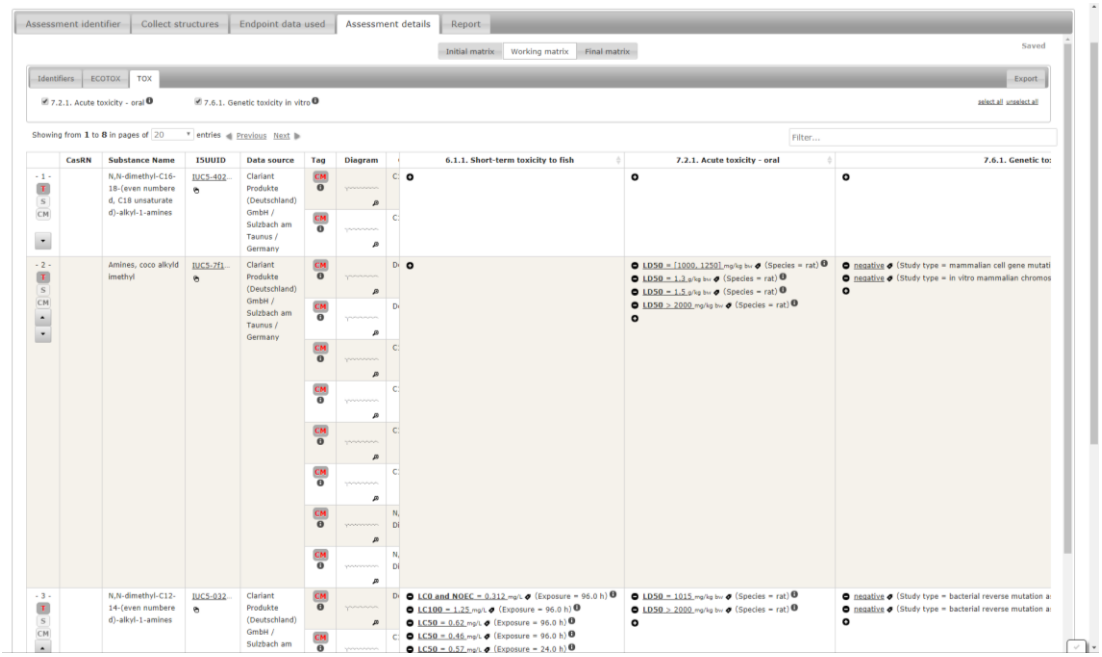
Showing 1 substance(s) (1 to 1) Previous Next

4.8 Water solubility (1)

Name	Endpoint	Result	Text	Guideline	Owner	Citation	Reliability	UUID

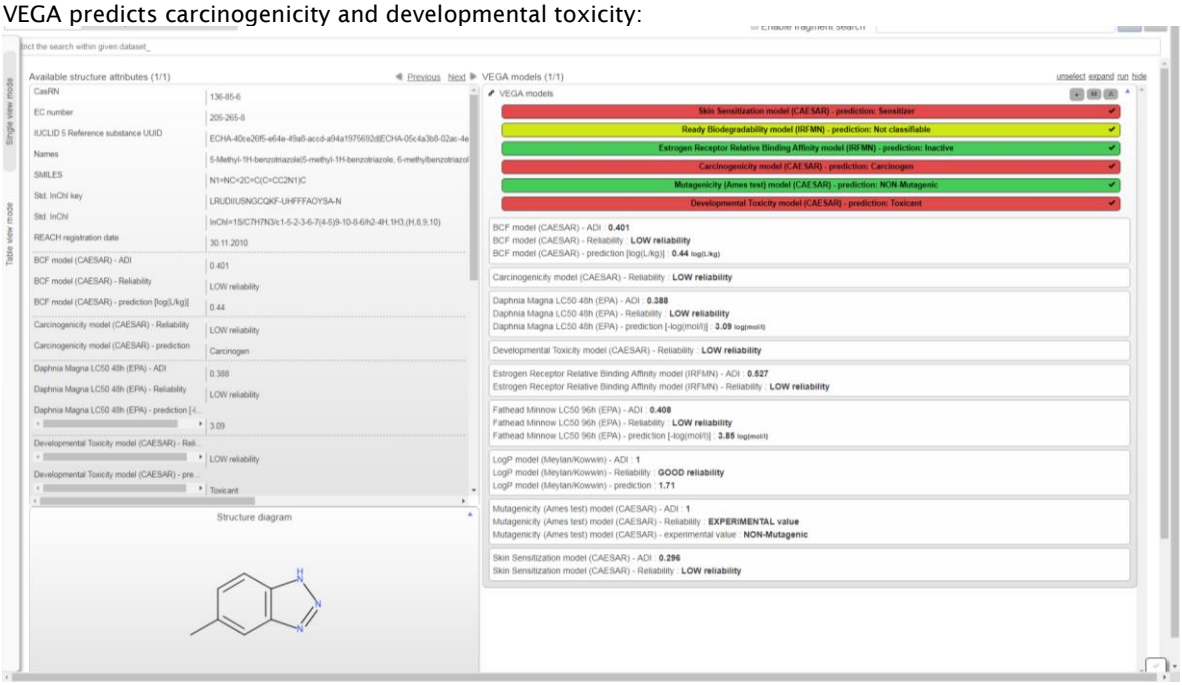
Read across

There are no data available to perform read across. Example of working matrix:



QSAR

VEGA predicts carcinogenicity and developmental toxicity:



ToxRead

Identity

N1=NC=2C=C(C=CC2N1)C

Measured data

Two QSAR models report absence of mutagenic activity.

Profiling

IRFMN alert n. 194 for NON-Mutagenicity, defined by the SMARTS: c1cn[n]n1

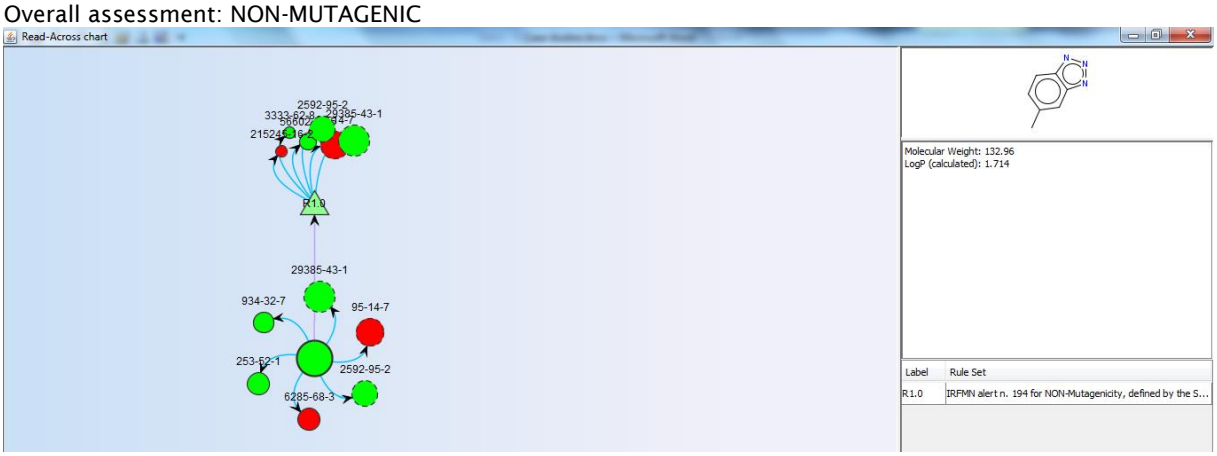
Experimental accuracy: 0.67
Fisher test p-value: 0.31587

Category definition
6 analogues:
29385-43-1 Similarity 0.976 Experimental activity: non mutagen
95-14-7 Similarity 0.935 Experimental activity: mutagen
2592-95-2 Similarity 0.892 Experimental activity: non mutagen
56602-32-5 Similarity 0.722 Experimental activity: non mutagen
3333-62-8 Similarity 0.605 Experimental activity: non mutagen
215245-16-2 Similarity 0.58 Experimental activity: mutagen

Read across
Read-Across assessment: Non-Mutagenic
Read-Across Mutagenic score = 0.21
Read-Across Non-Mutagenic score = 0.79

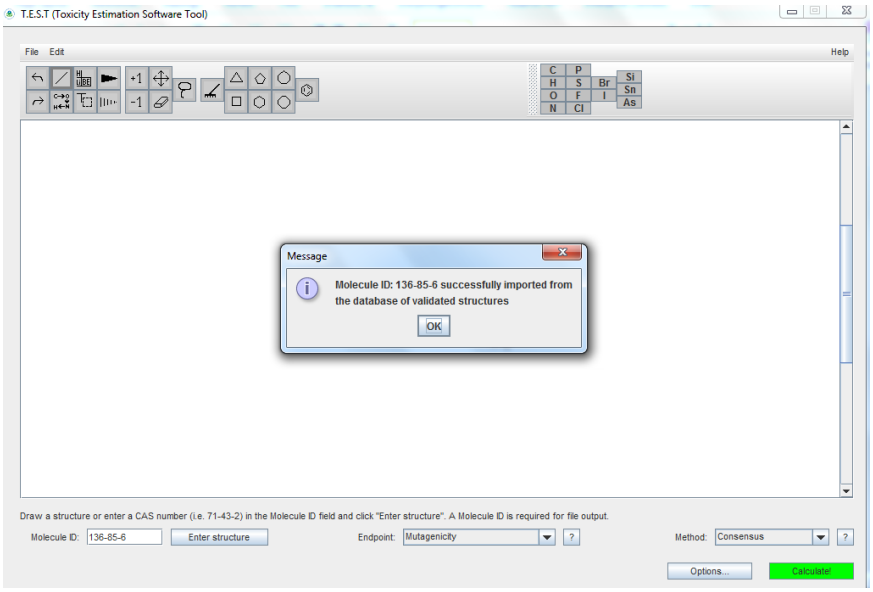
QSAR
QSAR consensus assessment: NON-Mutagenic (Consensus score: 0.55)
Predicted Consensus Mutagen activity = NON-Mutagenic
Consensus Score = 0.55
Model Caesar assessment = NON-Mutagenic (EXPERIMENTAL value)
Model ISS assessment = NON-Mutagenic (low reliability)
Model SarPy assessment = NON-Mutagenic (EXPERIMENTAL value)
Model KNN assessment = Mutagen (low reliability)

Overall assessment: NON-MUTAGENIC



T.E.S.T.

Identity
CAS number is present in database



Measured data

A negative test result for mutagenicity is reported.

Predicted Mutagenicity for 136-85-6 from Consensus method

Prediction results

Endpoint	Experimental value (CAS=136-85-6) Source: Toxicity Benchmark	Predicted value[a]
Mutagenicity value	0,00	0,70
Mutagenicity result	Mutagenicity Negative	Mutagenicity Positive

[a]Note: the test chemical was present in the training set. The prediction does not represent an external prediction.

Individual Predictions

Method	Predicted value
Hierarchical clustering	0,43
FDA	0,67
Nearest neighbor	1,00

Test chemical

QSAR

The consensus model predicts mutagenicity and absence of developmental toxicity. Similar chemicals were identified that were mutagenic as well. No chemicals with high similarity and measured developmental toxicity data were identified.

Predictions for the test chemical and for the most similar chemicals in the training set

If the predicted value matches the experimental values for similar chemicals in the training set (and the similar chemicals were predicted well), one has greater confidence in the predicted value.

Prediction statistics for similar chemicals

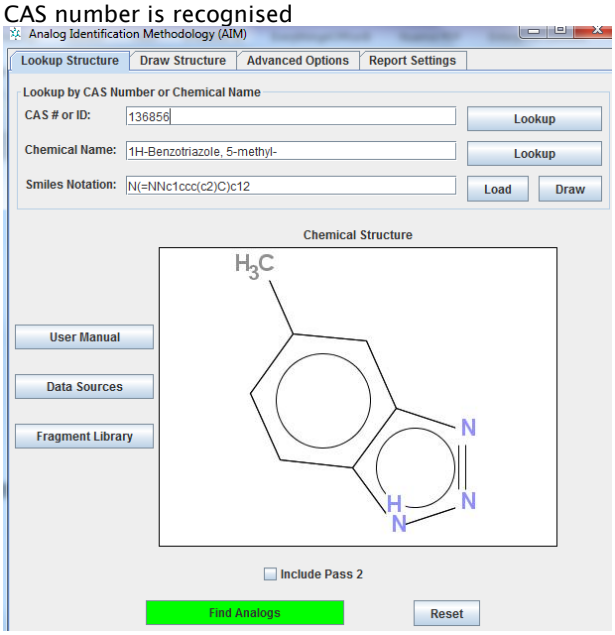
Concordance	Sensitivity	Specificity
0,80 (8 out of 10)	0,75 (3 out of 4)	0,83 (5 out of 6)

CAS	Structure	Similarity Coefficient	Experimental value	Predicted value
136-85-6 (test chemical)			0,00	0,70
95-14-7		0,95	1,00	0,42
2942-42-9		0,93	1,00	0,59
5401-94-5		0,92	1,00	0,79
114607-46-4		0,76	0,00	0,19

AIM

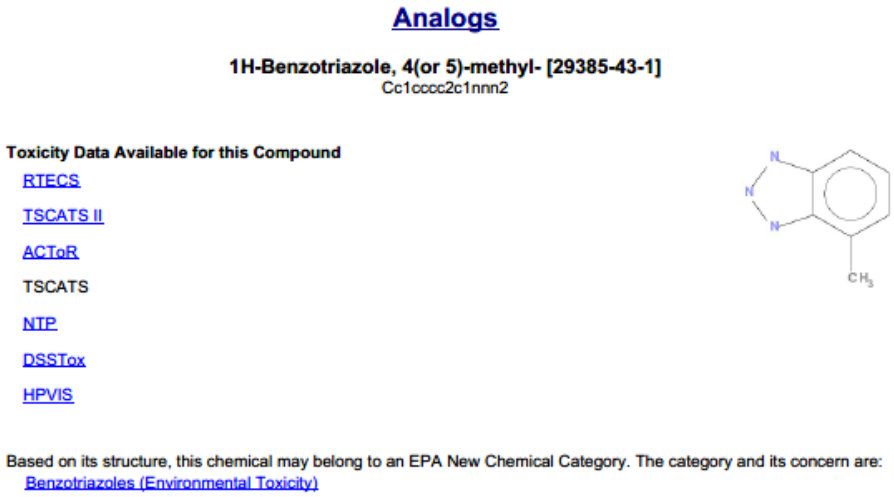
Identity

CAS number is recognised



Category definition

Pass 1 and 2 including 'Advanced options' yield one analogue with measured data: CAS 29385-43-1



Toxtree

Identity

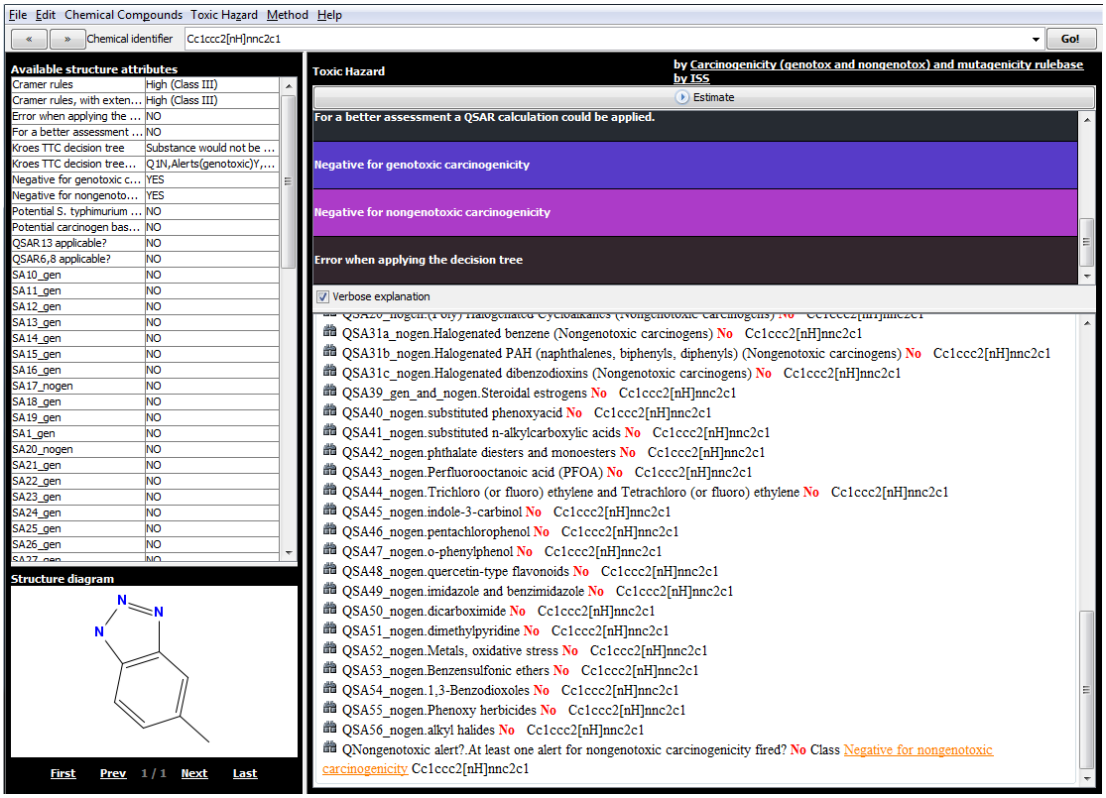
SMILES code entered

Profiling

Cramer class III

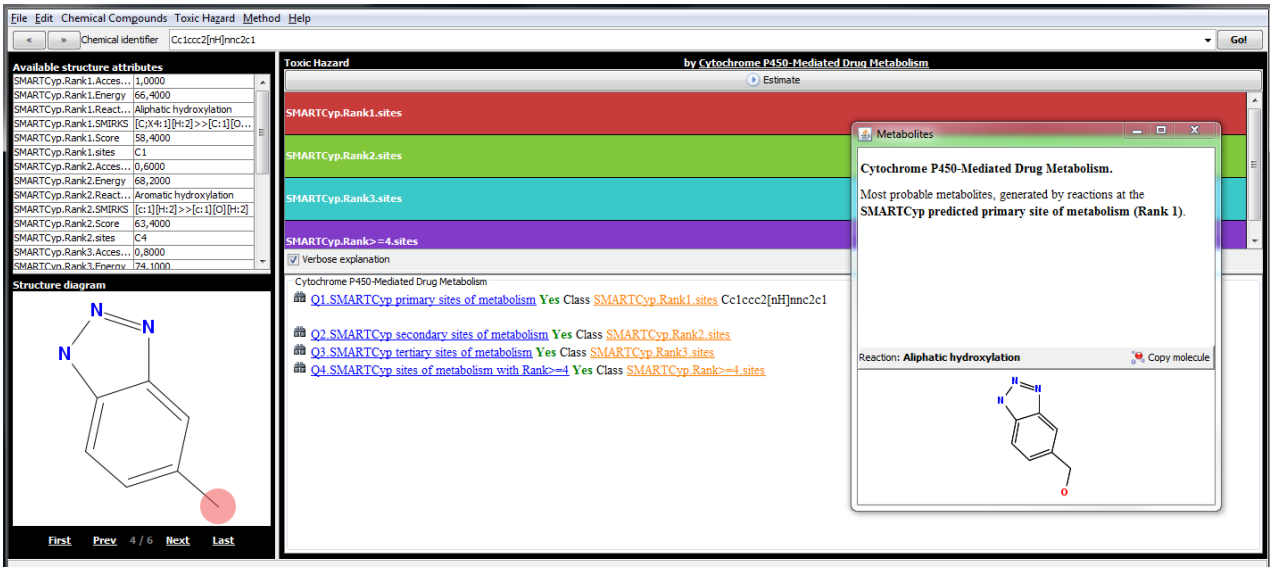
DNA binding alert (Michael acceptor)

No alerts for Ames mutagenicity; positive structural alerts for the micronucleus assay (H-acceptor-path3-H-acceptor)
Negative for (non-)genotoxic carcinogenicity



<u>Metabolites</u>

Four most likely metabolites:
OCc1ccc2NN=Nc2(c1)
CC1=CC(O)=C2NN=NC2(=C1)
OC=1C=C2NN=NC2(=CC=1C)
OC1=C2N=NNC2(=CC=C1C)



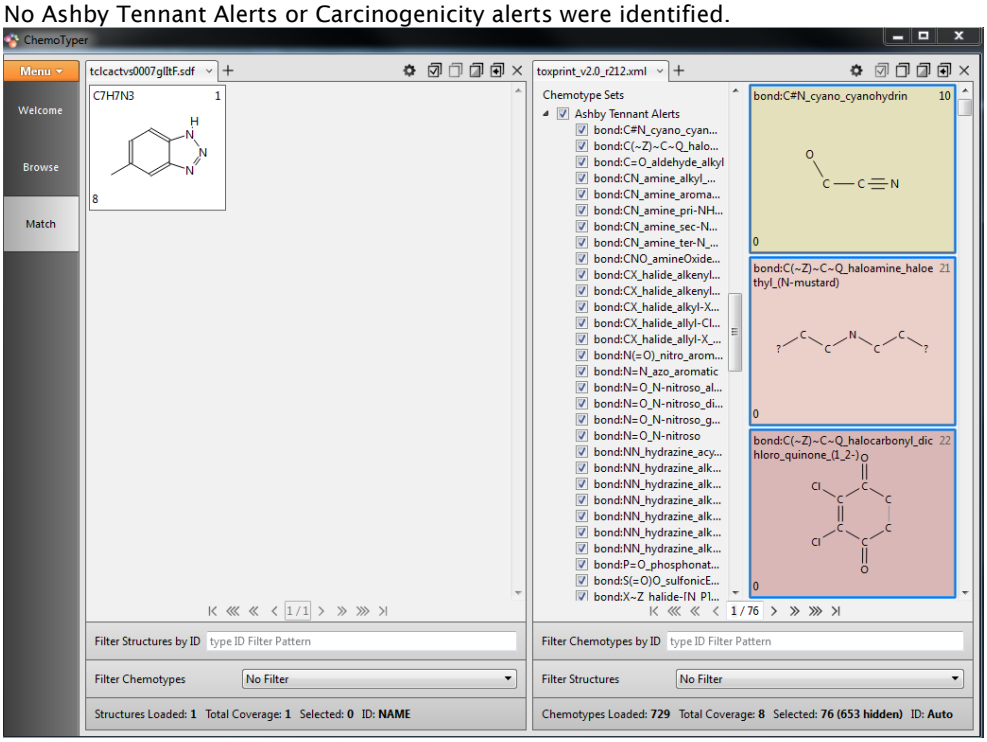
Chemotyper

Identity

Sdf file uploaded

Profiling

No Ashby Tennant Alerts or Carcinogenicity alerts were identified.

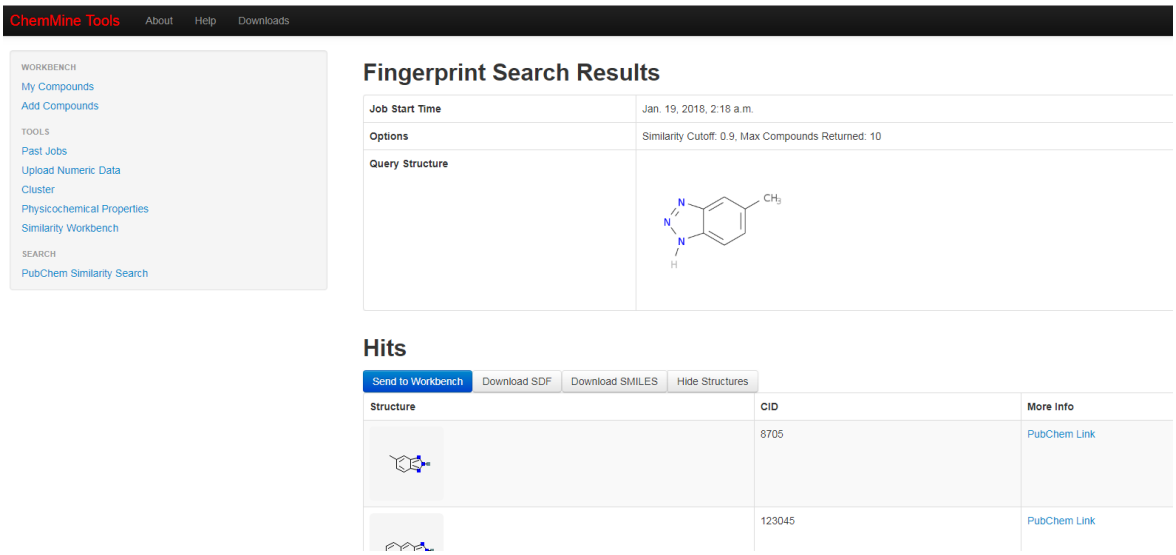


ChemMine

Category definition

Similarity Cutoff: 0.9, Max Compounds Returned: 10

CC1=CC2=N[NH]N=C2C=C1 [*OECD QSAR Toolbox CAS 136-85-6*]
[NH]1N=C2C=C3C=CC=CC3=CC2=N1 [*OECD QSAR Toolbox CAS 269-12-5*]
CC1=CC=CC2=N[NH]N=C12 [*OECD QSAR Toolbox CAS 29385-43-1*]
CC1=CC2=N[NH]N=C2C=C1C [*OECD QSAR Toolbox CAS 4184-79-6*]
FC(F)(F)C1=CC2=N[NH]N=C2C=C1 [*OECD QSAR Toolbox CAS -*]
CC1=CC2=N[Se]N=C2C=C1 [*OECD QSAR Toolbox CAS 1123-91-7*]
CCCN(CCC)CCC1=CC2=N[NH]N=C2C=C1 [*OECD QSAR Toolbox CAS -*]
[NH]1N=C2C=CC(=CC2=N1)C3=CC=CC=C3 [*OECD QSAR Toolbox CAS -*]
CC1=CC2=NSN=C2C=C1 [*OECD QSAR Toolbox CAS -*]
CC1=CC=C(C=C1)[N]2N=C3C=CC(=CC3=N2)C [*OECD QSAR Toolbox CAS -*]



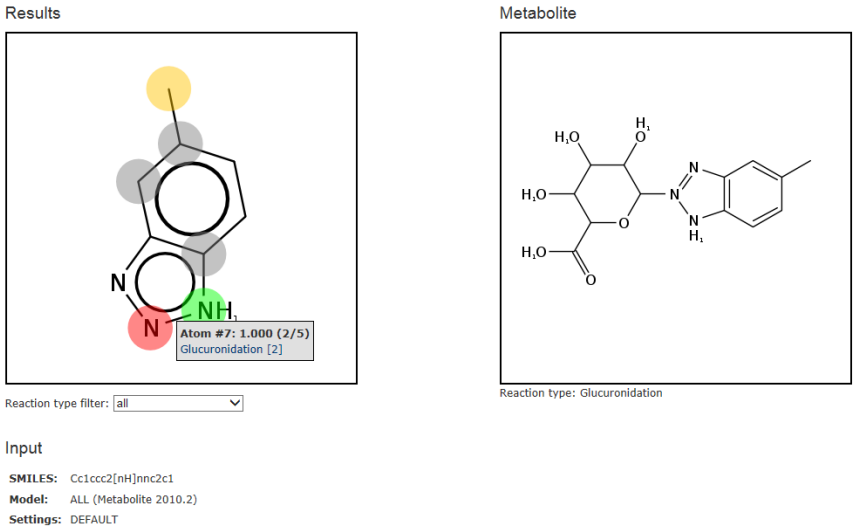
Metaprint2D-REACT

Identity

SMILES code entered

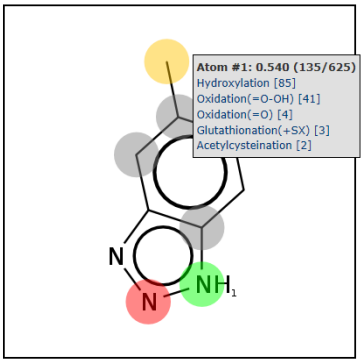
Metabolites

Fingerprint Matching default and Model ALL yields glucuronidation as the main metagbolic reaction.



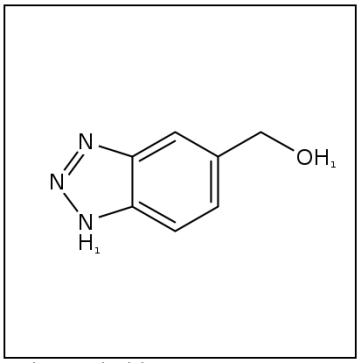
As the next probable site of metabolism several reactions are reported, of which hydroxylation is the most probable.

Results



Reaction type filter: all

Metabolite



Reaction type: Hydroxylation

Input

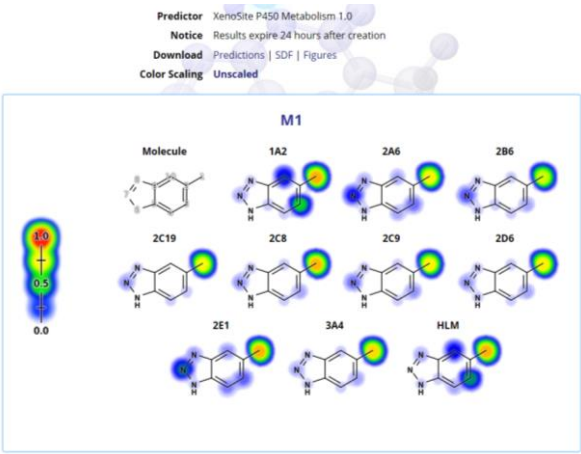
SMILES: Cc1ccc2[nH]nnc2c1
Model: ALL (Metabolite 2010.2)
Settings: DEFAULT

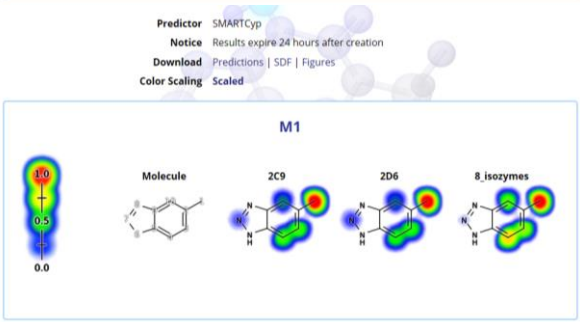
Xenosite

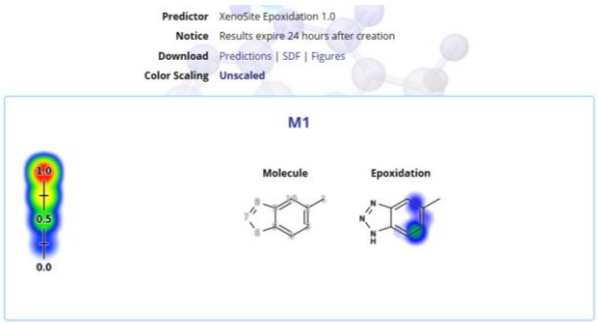
Identity
SMILES code entered

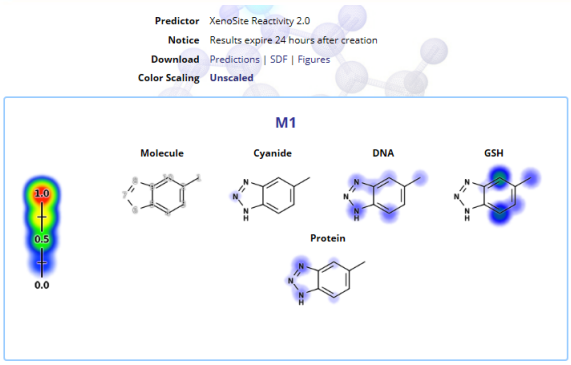
Profiling
DNA reactivity seems not likely.

Metabolites
The most probable reaction sites are shown below.





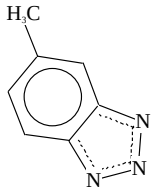




EPI Suite

Identity
CAS number entered
Profiling

EPI Suite Results For CAS 136-85-6



SMILES : n(nnc1ccc(c2)C)c12
CHEM : 1H-Benzotriazole, 5-methyl-
MOL FOR: C7 H7 N3
MOL WT : 133.15
---------- EPI SUMMARY (v4.11) ----------
Physical Property Inputs:
Log Kow (octanol-water): ------
Boiling Point (deg C) : ------
Melting Point (deg C) : ------
Vapor Pressure (mm Hg) : ------
Water Solubility (mg/L): ------
Henry LC (atm-m3/mole) : ------

Log Octanol-Water Partition Coef (SRC):
Log Kow (KOWWIN v1.68 estimate) = 1.71

Boiling Pt, Melting Pt, Vapor Pressure Estimations (MPBPVP v1.43):
Boiling Pt (deg C): 311.65 (Adapted Stein & Brown method)
Melting Pt (deg C): 97.46 (Mean or Weighted MP)
VP(mm Hg,25 deg C): 0.000314 (Modified Grain method)
VP (Pa, 25 deg C) : 0.0418 (Modified Grain method)
MP (exp database): 80-82 deg C
BP (exp database): 210-212 @ 12 mm Hg deg C
Subcooled liquid VP: 0.00105 mm Hg (25 deg C, Mod-Grain method)
: 0.14 Pa (25 deg C, Mod-Grain method)

Water Solubility Estimate from Log Kow (WSKOW v1.42):
Water Solubility at 25 deg C (mg/L): 3069
log Kow used: 1.71 (estimated)
no-melting pt equation used

Water Sol Estimate from Fragments:
Wat Sol (v1.01 est) = 10100 mg/L

ECOSAR Class Program (ECOSAR v1.11):
Class(es) found:
Benzotriazoles

Henrys Law Constant (25 deg C) [HENRYWIN v3.20]:
Bond Method : 1.62E-007 atm-m3/mole (1.64E-002 Pa-m3/mole)
Group Method: Incomplete
For Henry LC Comparison Purposes:
User-Entered Henry LC: not entered
Henrys LC [via VP/WSol estimate using User-Entered or Estimated values]:
HLC: 1.793E-008 atm-m3/mole (1.816E-003 Pa-m3/mole)
VP: 0.000314 mm Hg (source: MPBPVP)
WS: 3.07E+003 mg/L (source: WSKOWWIN)

Log Octanol-Air Partition Coefficient (25 deg C) [KOAWIN v1.10]:
Log Kow used: 1.71 (KowWin est)
Log Kaw used: -5.179 (HenryWin est)
Log Koa (KOAWIN v1.10 estimate): 6.889
Log Koa (experimental database): None

Probability of Rapid Biodegradation (BIOWIN v4.10):
Biowin1 (Linear Model) : 0.7388
Biowin2 (Non-Linear Model) : 0.8449
Expert Survey Biodegradation Results:
Biowin3 (Ultimate Survey Model): 2.8301 (weeks)
Biowin4 (Primary Survey Model) : 3.5871 (days-weeks)
MITI Biodegradation Probability:
Biowin5 (MITI Linear Model) : 0.3821
Biowin6 (MITI Non-Linear Model): 0.3428
Anaerobic Biodegradation Probability:
Biowin7 (Anaerobic Linear Model): 0.1780

Ready Biodegradability Prediction: NO

Hydrocarbon Biodegradation (BioHCwin v1.01):
Structure incompatible with current estimation method!

Sorption to aerosols (25 Dec C)[AEROWIN v1.00]:
Vapor pressure (liquid/subcooled): 0.14 Pa (0.00105 mm Hg)
Log Koa (Koawin est): 6.889
Kp (particle/gas partition coef. (m3/ug)):
Mackay model : 2.14E-005
Octanol/air (Koa) model: 1.9E-006
Fraction sorbed to airborne particulates (phi):
Junge-Pankow model : 0.000773
Mackay model : 0.00171
Octanol/air (Koa) model: 0.000152

Atmospheric Oxidation (25 deg C) [AopWin v1.92]:
Hydroxyl Radicals Reaction:
OVERALL OH Rate Constant = 2.7466 E-12 cm3/molecule-sec
Half-Life = 3.894 Days (12-hr day; 1.5E6 OH/cm3)
Half-Life = 46.731 Hrs
Ozone Reaction:
No Ozone Reaction Estimation
Fraction sorbed to airborne particulates (phi):
0.00124 (Junge-Pankow, Mackay avg)
0.000152 (Koa method)
Note: the sorbed fraction may be resistant to atmospheric oxidation

Soil Adsorption Coefficient (KOCWIN v2.00):
Koc : 84.91 L/kg (MCI method)
Log Koc: 1.929 (MCI method)
Koc : 87.87 L/kg (Kow method)
Log Koc: 1.944 (Kow method)

Aqueous Base/Acid-Catalyzed Hydrolysis (25 deg C) [HYDROWIN v2.00]:
Rate constants can NOT be estimated for this structure!

Bioaccumulation Estimates (BCFBAF v3.01):
Log BCF from regression-based method = 0.798 (BCF = 6.281 L/kg wet-wt)
Log Biotransformation Half-life (HL) = -0.7170 days (HL = 0.1919 days)
Log BCF Arnot-Gobas method (upper trophic) = 0.742 (BCF = 5.517)
Log BAF Arnot-Gobas method (upper trophic) = 0.742 (BAF = 5.517)
log Kow used: 1.71 (estimated)

Volatilization from Water:
Henry LC: 1.62E-007 atm-m3/mole (estimated by Bond SAR Method)
Half-Life from Model River: 4171 hours (173.8 days)
Half-Life from Model Lake : 4.56E+004 hours (1900 days)

Removal In Wastewater Treatment:
Total removal: 2.06 percent
Total biodegradation: 0.09 percent

Total sludge adsorption: 1.96 percent
Total to Air: 0.01 percent
(using 10000 hr Bio P,A,S)

Level III Fugacity Model:

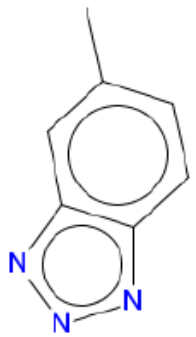
Mass Amount (percent)	Half-Life (hr)	Emissions (kg/hr)
Air 1.7	93.5	1000
Water 23.2	360	1000
Soil 75	720	1000
Sediment 0.122	3.24e+003	0

Persistence Time: 589 hr

VEGA

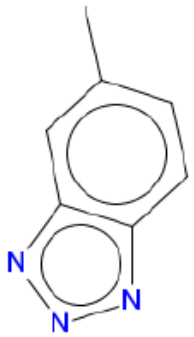
Identity
SMILES code entered

Measured data / QSAR
All predictions are listed below:



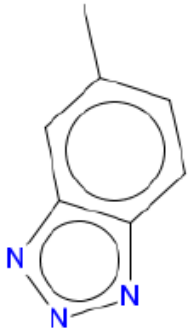
Prediction:

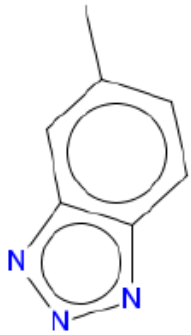
Prediction is NON-Mutagenic with a consensus score of 1, based on 2 experimental values.



EXPERIMENTAL DATA

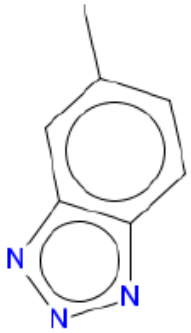
Experimental value is NON-Mutagenic. Model prediction is NON-Mutagenic (good reliability).

	**EXPERIMENTAL DATA** **Experimental value is NON-Mutagenic. Model prediction is Possible NON-Mutagenic (good reliability).**



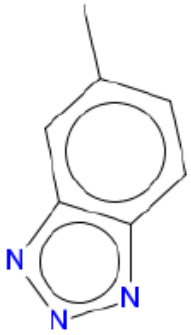
Prediction: Reliability:

Prediction is NON-Mutagenic, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 infrequent fragments found)



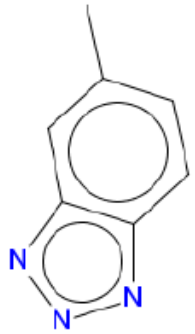
Prediction: Reliability:

Prediction is Mutagen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- accuracy of prediction for similar molecules found in the training set is not adequate
- some similar molecules found in the training set have experimental values that disagree with the predicted value



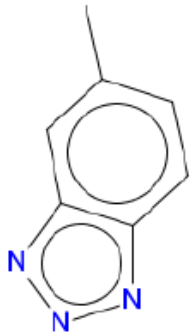
Prediction: Reliability:

Prediction is Carcinogen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 infrequent fragments found)
- model class assignment is uncertain
- predicted value disagrees with experimental values of training set compounds laying in the same neuron



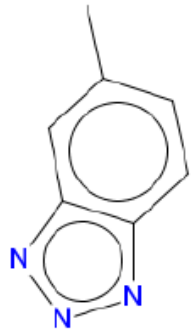
Prediction: Reliability:

Prediction is NON-Carcinogen, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- only moderately similar compounds with known experimental value in the training set have been found
- accuracy of prediction for similar molecules found in the training set is not adequate
- some similar molecules found in the training set have experimental values that disagree with the predicted value
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 infrequent fragments found)



Prediction: Reliability:

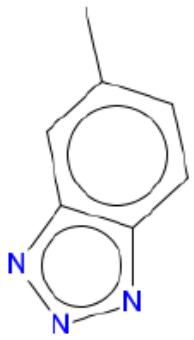
Prediction is Possible NON-Carcinogen, the result appears reliable. Anyhow, you should check it through the evaluation of the information given in the following sections. Anyway some issues could be not optimal:
- some similar molecules found in the training set have experimental values that disagree with the predicted value



Prediction: Reliability:

Prediction is Possible NON-Carcinogen, but the result shows some critical aspects, which require to be checked:
- accuracy of prediction for similar molecules found in the training set is not adequate
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 infrequent fragments found)

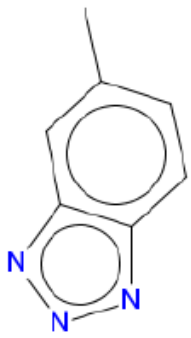
Developmental toxicity:



Prediction: Reliability:

Prediction is Toxicant, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- only moderately similar compounds with known experimental value in the training set have been found
- accuracy of prediction for similar molecules found in the training set is not adequate
- similar molecules found in the training set have experimental values that disagree with the predicted value
- some atom centered fragments of the compound have not been found in the compounds of the training set or are rare fragments (1 infrequent fragments found)

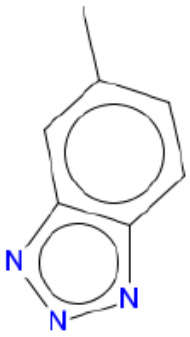
Developmental/reproductive toxicity:



Prediction: Reliability:

Prediction is NON-Toxicant, but the result may be not reliable. A check of the information given in the following section should be done, paying particular attention to the following issues:
- only moderately similar compounds with known experimental value in the training set have been found
- similar molecules found in the training set have experimental values that disagree with the predicted value

Hepatotoxicity:



Prediction: Reliability:

Prediction is Toxic, but the result shows some critical aspects, which require to be checked:
- only moderately similar compounds with known experimental value in the training set have been found
- accuracy of prediction for similar molecules found in the training set is not adequate
The following relevant fragments have been found: Hepatotoxicity toxic alert no. 40

Attachment IV

Data collection sheet

A default sheet that can be used to summarize and compare the information retrieved using multiple data sources and tools is presented on the next page.

	Target chemical	Analogues					
Name **CAS** **SMILES**							
Metabolites							
Observed (O) / Predicted (P) mammalian CYP450 rat *in vivo* rat S9 microbial environmental							
Measured data							
Systemic toxicity Genotoxicity Carcinogenicity Reproductive & developmental toxicity							
Structural alerts							
Cramer class Systemic toxicity Genotoxicity *in vitro* Genotoxicity *in vivo* Carcinogenicity Reproductive toxicity Developmental toxicity							
Category formation							
Rationale:							
Structural similarity Functional groups Structural alerts Physico-chemical properties: • MW • MP • BP • VP • Log Kow • Water solubility • Lipinski rule	*Score:*						
Read across							
Endpoint: Endpoint:		*Justification:* *Justification:*					
QSAR							
Endpoint: Endpoint:							