Computational Methods in Support of Chemical Risk Assessment

Mojca Fuart Gatnik

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

This research programme was carried out in collaboration with the Joint Research Centre of the European Commission

June 2016

Acknowledgements

Immense gratitude goes to my Director of Studies and academic supervisor, Mark Cronin. I thank Mark for accepting me at the Liverpool John Moores University and being a truly dedicated mentor, providing me with academic support and encouragement. I am particularly indebted to him for giving me the unique opportunity to present my work at EUROTOX 2015.

The same profound gratitude goes to my direct supervisor, Andrew Worth, who gave me the wonderful opportunity to obtain the Grant for a PhD in the field of Computational Toxicology at the Joint Research Centre in Ispra (Italy). And not only for that but also for his dedicated assistance and brain-storming discussions all the way through the period at the JRC and afterwards. I have very fond memories of my time there.

A very special thank you goes to my academic supervisor, Judith Madden, for her constant support and for so generously dedicating her time during my stay in Liverpool.

I am wholeheartedly grateful to Judy, Mark and Andrew, who believed in me and my work and encouraged me throughout a difficult period, as without their support and understanding the way to this moment would not have been possible.

I would also like to extend my gratitude to everyone from the QSAR Laboratory at LJMU for their kindness and generosity, which always made my visits to Liverpool a fulfilling experience.

I am hugely appreciative of my work colleague, Rositsa Serafimova, especially for sharing her knowledge of chemistry and for being available for discussion at any time.

It was a great pleasure and an inspiring experience to exchange thoughts and ideas with people from different backgrounds. A special mention goes to Chihae Yang for the critical discussion on possibilities regarding the extrapolation of toxicological data for biocides and to Leon van der Wal for his critical opinion on the application of TTC.

I would like to thank Susan Owen Molinario for becoming my friend when I moved to Italy and for encouraging me during my advanced English classes.

Finally, but by no means least, I would like to thank my husband Flavio and son Lovro for being unbelievably patient and supportive and also to my parents for their unselfish share of childcare in crucial moments.

Abstract

Chemical risk assessment for human health effects is performed in order to establish safe exposure levels of chemicals to which individuals are exposed. The process of risk assessment traditionally involves the generation of toxicological studies from which health based guidance values are derived for a specific chemical. For low level exposures to chemicals, where there are no or limited chemical specific toxicity data, the application of the Threshold of Toxicological Concern (TTC) approach may estimate whether the exposure levels can be considered safe. The TTC approach has recently gained increasing interest as new requirements, under different regulatory frameworks, emerge for the safety assessment of chemicals and to assess chemicals for which testing is not routinely required.

The application of TTC relies heavily on computational (*in silico*) methods. *In silico* tools are computer implemented models that, based on commonalities in the toxicity of "similar" chemical structures, may predict hazard. *In silico* methods are rapidly evolving and gaining importance within the context of Integrated Approaches to Testing and Assessment (IATA) and their acceptance for regulatory purposes is expanding. The work presented in this thesis has focused on the use and applicability of a wide range of computational approaches to assist in the application of the TTC concept.

In the TTC approach, the identification of genotoxic chemicals is a primary requirement. *In silico* approaches apply expert knowledge and/or statistical methods to either predict genotoxicity or to identify structural alerts associated with it. This thesis focused, in part, on a group of important environmental pollutants, nitrobenzenes, to assess the applicability of *in silico* tools to predict genotoxicity. For this purpose a dataset containing 252 nitrobenzenes including Ames test results was compiled. Based on these test results a case study for sodium nitro-guaiacolate, a pesticide active substance, was developed. The case study demonstrated that (Q)SAR and a category approach incorporating read-across, are applicable for the prediction of genotoxicity and supports their use within a weight of evidence approach.

Another aspect of the TTC approach is the evaluation of repeat dose, non-cancer endpoints. For that purpose chemicals are separated into groups related to three levels of concern based on the Cramer classification. For each level, namely the Cramer Classes (I, II and III), a safe exposure level has been established. Therefore, as interest to apply TTC expands to new groups of chemicals, the reliability and conservativeness of the established thresholds relative to Cramer Classes for the new chemistries must be established. In this thesis the TTC approach was evaluated for 385 cosmetic ingredients, 77 biocides and 102 compounds classified as reproductive and developmental toxicants. To support the evaluation at different levels, chemical datasets containing toxicological data were utilised and computational tools were applied to compare datasets. The results indicated, that the historical "Munro" dataset is broadly representative for cosmetics and biocides. In addition, that the threshold levels for Cramer Class III are within the range of Munro's threshold further supports the validity of the TTC approach and its conservativeness for the groups of chemicals analysed. Cramer Class I thresholds were found to be valid only for classified developmental and reproductive toxicants. The results also supported the validity of the classification of chemicals into Cramer class III.

It is foreseen that the TTC approach will gain increasing acceptance in the risk assessment of different groups of chemicals. Therefore it is emphasised that the future work should focus on the identification of the limitations of the application of TTC, including the identification of groups of chemicals to which TTC cannot be applied, the expansion of the underlying toxicological datasets, and the development of tools to support the application of TTC so that is transparent and acceptable for regulatory purposes.

Abbreviations

AChE	Acetylcholinesterase
AO	Adverse Outcome
AOP	Adverse Outcome Pathway
BIT	Benzisothiazolinone
BKC	Benzalkonium chloride
BMD	Benchmark dose
BMDL10	Lower confidence limit on a benchmark dose giving 10% increase in tumours
BPR	Biocidal Products Regulation
CAS	Chemical Abstracts Service
CCRIS	Chemical Carcinogenesis Research Information System
CDF	cumulative distribution function
CLP	Classification, labelling and packaging of substances and mixtures
CMR	Carcinogenicity, Mutagenicity and Reproductive substance
COC	Cohort of Concern
COSMOS	Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of COSMetics to Optimise Safety
CPDB	Carcinogenicity Potency Database
CPNP	Cosmetic Products Notification Portal
DAR	Draft Assessment Report
DB	Database
DDAC	Dialkyldimethylammonium chloride
DfW	Derek for Windows
DG SANTE	Directorate General for Health and Food Safety
DNA	Deoxyribonucleic acid
DPD	Dangerous Preparations Directive
DSD	Dangerous Substance Directive
DSSTox	Distributed Structure-Searchable Toxicity
DST	Dermal Sensitisation Threshold
EAFUS	"Everything" Added to Food in the United States
EBW	Exposure Based Waiving
EC3	Effective Concentration -threefold
ECHA	European Chemicals Agency
ED	Endocrine disruption
EDC	Endocrine Disrupting Chemicals
EFSA	European Food Safety Authority
EMA	European Medicines Agency
EPA	United States Environmental Protection Agency
EPAA	European Partnership for Alternative Approaches to Animal Testing
EU	European Union
FAO	Food and Agriculture Organization
FDA	US Food and Drug Administration
GRAS	Generally Recognized As Safe
HBGV	Health Based Guidance Values
HESI	Health and Environmental Sciences Institute
IATA	Integrated Approach to Testing and Assessment
ICCR	International Cooperation on Cosmetics Regulation
IHCP	Institute for Health and Consumer Protection
ILSI	International Life Sciences Institute
IPS	Integrated Prediction System

ISS	Istituto Superiore di Sanità
ITEM	Institute for Toxicology and Experimental Medicine
ITS	Integrating testing strategy
IUPAC	International Union of Pure and Applied Chemistry
JECFA	Joint FAO/WHO Expert Committee on Food Additives
JMPR	Joint Meeting on Pesticide Residues
JRC	Joint Research Centre
LEL	Lowest Effect Level
LOAEL	Lowest Observed Adverse Effect level
LOEL	Lowest Observed Effect level
MoA	Mode of Action
MOE	Margin of Exposure
NCCT	United States National Centre for Computational Toxicology
NIEHS	National Institute of Environmental Health Sciences
NL	The Netherlands
NO(A)EL	No Observed (Adverse) Effect Level
NTP	National Toxicology Program
OECD	Organisation for Economic Co-operation and Development
OPP	US EPA Office of Pesticide Programs
PAFA	Priority-based Assessment of Food Additives
POD	Point of Departure
PPP	Plant Protection Products
PSA	Molecular polar surface area
(Q)SAR	QSARs and SARs are collectively referred as (Q)SARs
RAR	Risk Assessment Report
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
RIVM	Dutch National Institute for Public Health and the Environment
RMS	Rapporteur Member State
RN	Registry Number
ROC	Receiver Operating Characteristics
RTECS	Registry of Toxic Effects on Chemical Substances
SA	Structural Alert
SAR	Structure-Activity Relationships
SCCP	Scientific Committee on Consumer Products
SCCS	Scientific Committee on Consumer Safety
SCENIHR	Scientific Committee on Emerging and Newly Identified Health Risks
SCHER	Scientific Committee on Health and Environmental Risks
SEURAT-1	Safety Evaluation Ultimately Replacing Animal Testing
SMILE	Simplified Molecular Input Line Entry System
TD	Tolerated Dose
TNT	Trinitrotoluene
ToR	Threshold of Regulation
TPSA	Topological Polar Surface Area
TTC	Threshold of Toxicological Concern
UF	Uncertainty Factors
US	United States of America
WHO	World Health Organization
WoE	Weight of Evidence approach

Contents

1 Introduction

1.1 Risk Assessment and Standard Toxicity Testing

Risk assessment is a key element of the control of chemicals. It attempts to anticipate or predict possible future harm to human health or the environment following exposure to one or more substance. As such, the traditional process of risk assessment encompasses hazard identification, hazard characterisation, exposure assessment, and risk characterisation. Hazard identification and characterisation are defined by toxicological testing, from which potential adverse effects and mechanisms of action can be identified. In this step, the relationships between the dose, route, duration and timing of exposure are assessed with all related uncertainties. Within the exposure assessment, the actual level of exposure of an individual to a toxicant is evaluated based on frequency, duration and levels of contaminants. As the final step in risk assessment, risk characterisation integrates conclusions from the previous steps. The overall aim of the process of risk assessment is the identification of safe levels of exposure according to which risk mitigation measures are applied, if necessary. While all chemicals have an inherent capacity to cause adverse effects (i.e. a hazard), not every exposure causes harm. The likelihood of harm due to exposure distinguishes risk from hazard (van Leeuwen & Vermeire, 2007).

1.2 Threshold of Toxicological Concern in the Context of Risk Assessment

The Threshold of Toxicological Concern (TTC) is a pragmatic risk assessment tool that is based on the principle of establishing a human exposure threshold value for all chemicals, below which there is low probability of an appreciable risk to human health (Barlow, 2005; Koster et al., 2011; Kroes, Kleiner, & Renwick, 2005). The TTC can be identified for chemicals present at low concentrations in the diet, for those with unknown toxicity, on the basis of their chemical structure. As such, it can be used for preliminary risk characterisation, by comparing potential exposure and the TTC value, based on the substance's structure. This could avoid the need for further detailed assessment of chemicals to which humans are exposed at low levels. It can be also used to set priorities in toxicity testing. The TTC approach can be applied in cases when experimental toxicity data are not available and negligible exposure is expected, but reliable

exposure assessment is required for that substance. Moreover, the TTC approach is not intended to be a substitute in circumstances where a detailed toxicological profile of a substance is required by regulatory framework (e.g. information on the pesticide active ingredients). The application of the TTC principle in the risk assessment process is presented in Figure 1.1 (adapted from (Kroes et al., 2005)).

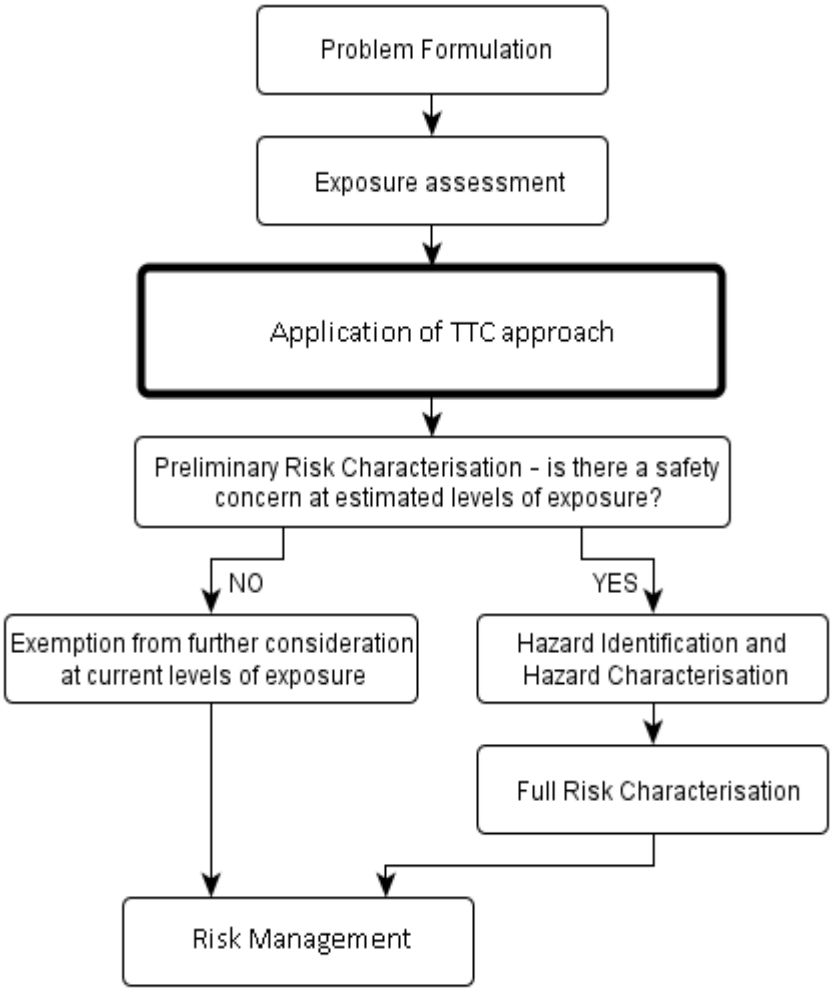


Figure 1.1. Application of the TTC principle in the Risk Assessment Paradigm (modified from Kroes, 2005).

1.2.1 TTC Values

A general approach for setting safe exposure limits associated with negligible risk to human health for toxicants includes the derivation of an experimentally safe dose (Hayes, 2008). The

experimentally safe dose for the application of TTC to carcinogenic chemicals was derived based on the analysis of distributions of experimentally derived TD50 for cancer endpoint to which no further uncertainty factors have been applied. In this case TD50 represents a median toxic dose and is a measure of carcinogenic potency, representing a dose in mg/kg body wt/day which would induce tumour(s) in half of the tested animals at the end of their normal lifespan. A lower value indicates a more potent carcinogen. In the case of non-cancer endpoint, the virtually safe dose is derived based on distributions of experimentally derived No Observed (Adverse) Effect Level (NO(A)EL) or values which have been obtained using the application of uncertainty factors.

The threshold for (genotoxic and non-genotoxic) carcinogenic effects was identified based on the analysis of carcinogenic potencies, which were log normally distributed by several authors including (Cheeseman, Machuga, & Bailey, 1999; Munro, 1990; Rulis, 1986). The threshold level of 0.5 ppb, equivalent to an intake of 1.5 μg/day, which is considered to be protective for known and unknown carcinogens, was established. The level of risk was extrapolated through a linear extrapolation of TD50 values, for compounds originating from the Carcinogenic Potency Database (CPDB), to a one in a million excess cancer risk. Linear extrapolation to a probability of 1 in 1,000,000 (i.e., the accepted lifetime risk level used) was achieved by simply dividing the TD50 by 500,000. The approach of linear extrapolation is considered conservative as it assumes that biological responses, in the process of a generation of a tumour, are linear along the range of extrapolation (Kroes et al., 2005). Separately, a group of chemicals with risk higher than one in a million even at exposure below the dietary threshold exposure was identified. Some of these compounds, the Cohort of Concern (COC), were excluded from the application of TTC, while for some chemicals containing a structural alert for genotoxicity or having a positive Ames Test, a lower threshold was set (Cheeseman et al., 1999; Kroes et al., 2004). As an indication of chemicals to which the TTC should not be applied, a COC includes three groups of genotoxic carcinogenic compounds (aflatoxin-like compounds, azoxy compounds, nitroso compounds) and two groups of high potency carcinogens acting through a hormone modulation activity mode of action (e .g. TCDD, steroids).

For non-cancer endpoints, the threshold values relative for each Cramer hazard class (defined by the work of Cramer et al (Cramer & Ford, 1978) as a structure-based decision tree, which is further explained in section 1.2.2.) were calculated from the derived 5th percentile NOEL, which was obtained from the analysis of a log normal distribution of NOEL values. The NOEL implies any adverse or non-adverse effect related to the exposure to a chemical where the NOAEL is the highest experimental dose below which no statistically, or biologically significant, increase in toxic effects was observed in the tested group; it is usually derived from a dose response analysis. The NOEL is also referred as a point of departure (POD) to derive health based guidance values (HBGV). The NO(A)EL approach is traditionally applied to derive safety levels for non-cancer endpoints assuming that a threshold exists in the dose response curve.

In order to account for uncertainties in the derived value, uncertainty factors (UF) are applied to define the final HBGV. Several categories of uncertainty factors have been developed to account for human variability, to extrapolate from animal to human data, to extrapolate from subchronic to chronic studies and from a LOAEL study to NOAEL, as well as an uncertainty factor to account for limited data for a chemical.

A set of threshold values, reported in Table 1.1, has been derived and has been applied since then (Cheeseman et al., 1999; Kroes et al., 2004; Munro, Ford, Kennepohl, & Sprenger, 1996; Rulis, 1986).

Table 1.1. Applicable threshold values

Type of threshold	**TTC value µg/person per day**	**TTC value µg/kg bw per day**
Structural alert for genotoxicity	0.15	0.0025
Structural alert for non – genotoxic carcinogenicity	1.5	0.025
Structural alert for Acetylcholinesterase- AChE inhibition (organophosphates and carbamates)	18	0.3
Cramer Class III	90	1.5
Cramer Class II	540	9.0
Cramer Class I	1800	30

The Margin of Exposure (MOE), an alternative methodology to assess the level of concern, was introduced in 2005 in the European Union (EU) for the assessment of the safety of carcinogenic and genotoxic compounds that are not intentionally added to products, but which may be present in low quantities (EFSA, 2005). In a similar manner to the TTC approach, the MOE represents a ratio between a dose descriptor (NOEL, BMD) and the expected exposure level in humans. In the derivation of MOE, a dose descriptor is determined based on dose response modelling to set a POD representing a concentration at which hazard related changes are low, but measurable. It is accepted that a POD in the case of genotoxic carcinogens should be a Benchmark dose (BMD), $BMDL_{10}$ (the lower confidence limit on a benchmark dose giving 10% increase in tumours) (EFSA, 2005). The higher the MOE value, the lower is the level of concern from exposure to a chemical. MOE, therefore, provides a method to compare the level of concern for different compounds and to prioritise risk management for unavoidable exposures to genotoxic and carcinogenic contaminates, allowing for decisions to be made on acceptable exposures.

1.2.2 The Cramer Decision Tree

The Cramer decision tree approach was applied to rank chemicals according to their expected hazard. This approach uses knowledge of structure-activity relationships (SARs), metabolism, chemical reactivity, human exposure levels and other relevant information (Cramer & Ford, 1978). The decision tree consists of 33 yes/no questions, leading to the final classification of a chemical into one of three classes reflecting the presumption of low, moderate and high toxicity. As a result, substances are classified into one of three classes:

- Class I (Low) contains substances of simple chemical structure with known metabolic pathways and innocuous end products which suggest a low order of oral toxicity.

- Class II (Intermediate) contains substances that possess structures that are less innocuous than those in Class I, but they do not contain structural features that are suggestive of toxicity such as those in Class III.

- Class III (High) contains substances with a chemical structure that permits no strong initial impression of safety and may even suggest a significant toxicity.

The Cramer scheme is applicable to organic molecules and their salts, while polymers, oligomers and inorganics cannot be classified by the decision tree.

A TTC decision tree process, depicted in Figure 1.2, was proposed by Kroes et al. (Kroes et al., 2004) to place the chemical under study into one of five levels of TTC ranging from 0.15 µg/day for potential genotoxic carcinogens to 1800 µg/day for non-cancer chemicals considered to be of low toxicity. A group of very potent carcinogens were excluded from the TTC approach; for these chemicals, substance-specific data are required to complete the risk assessment.

To support risk assessors in the application of the Kroes TTC decision tree, Cramer rules have been coded into several pieces of software. One such piece of software was the result of an initiative of the Joint Research Centre of the European Commission. A current version of Toxtree-v2.6.13. is available for download through Sourceforge (http://toxtree.sourceforge.net/download.html).

1.2.3 Historical Overview of TTC

The postulation by Paracelsus that "*the dose makes the poison*", was put into the context of toxicologically insignificant exposure levels to chemicals of unknown toxicity, due to an increasing demand for toxicity testing, by Frawley in 1967 (Frawley, 1967). As a result, the cut-off value of 1.5µg/day was then accepted by the US (United States) Food and Drug Administration (FDA) as a Threshold of Regulation (ToR), which meant that no further testing was required for substances migrating from packaging into food below this level of exposure.

Over time, the TTC concept has been developed and refined by several authors (Cheeseman et al., 1999; Kroes et al., 2000, 2004; Munro et al., 1996; Munro, 1990) with the aim of providing a tiered approach. This approach should act as a guide on how, and when, the TTC principle could be applied, as a preliminary step, in the safety evaluation of chemicals. A set of threshold levels, based on chronic systemic toxicity studies after oral exposure, was derived. These

thresholds are meant to represent an exemption from testing, as no risk is expected at levels below the thresholds. Based on these studies, the currently accepted TTC values are reported in Table 1.1.

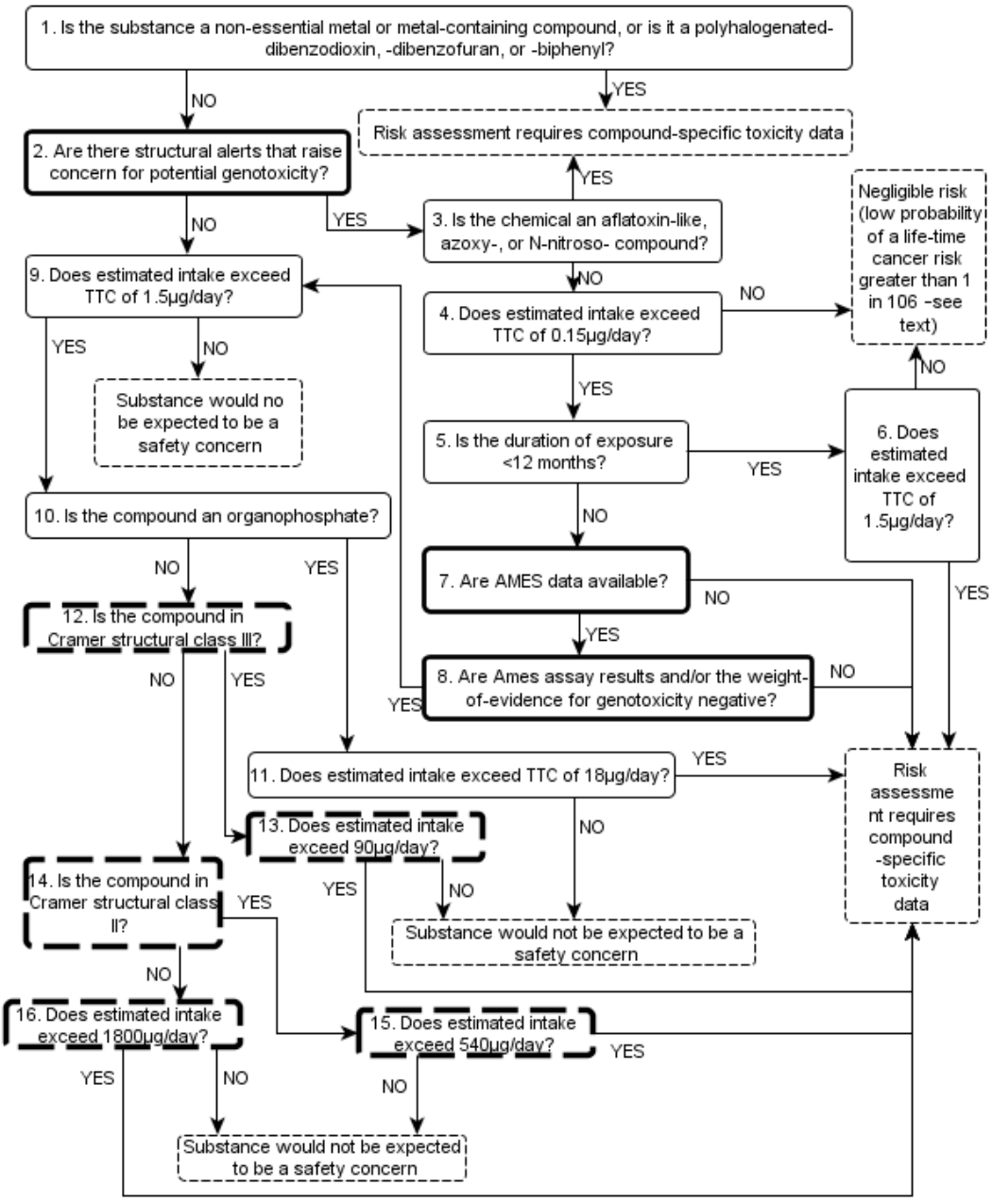


Figure 1.2. Decision tree for risk prioritisation of chemicals as proposed by Kroes et al., 2004 including the extension proposed by Felter et al., 2009 (Felter et al., 2009). Bolded (solid and dashed lines) nodes are relevant for the study presented in Chapter 3

Although primarily developed for migrant packaging material chemicals present at low levels in food (Cheeseman, 2005; Munro, Hlywka, & Kennepohl, 2002), the TTC concept has also

been successfully applied to food additives (Munro, Kennepohl, & Kroes, 1999; Renwick, 2004), and to genotoxic impurities in pharmaceutical preparations (EMEA, 2006). It has also been proposed for ingredients used in personal care products (Kroes et al., 2007) and pesticide metabolites (EFSA, 2012a).

Since then, especially in recent years, due to increasing regulatory requirements and promotion of alternative methods to animal testing under a number of pieces of European legislation (e.g. plant protection products (PPP), biocidal products (BPR), industrial chemicals (REACH) and cosmetic products) most efforts have been oriented towards the evaluation of the validity and applicability of structural thresholds to assess specific endpoints or specific chemical classes. The European Food Safety Authority (EFSA) extensively evaluated the utility of TTC for its application to areas under EFSA's remit, such as pesticides metabolites (EFSA, 2012a).

The application of thresholds for developmental and reproductive toxicity was evaluated in several studies by Bernauer et al. (Bernauer, Heinemeyer, Heinrich-Hirsch, Ulbrich, & Gundert-Remy, 2008), van Ravenzwaay et al. (van Ravenzwaay, Dammann, Buesen, Flick, & Schneider, 2012), Laufersweiler et al. (Laufersweiler et al., 2012) and Stanard et al. (Stanard, Dolan, Hanneman, Legare, & Bercu, 2015). The conclusions of these reports are discussed in more depth in Chapter 6 in this thesis.

TTC is primarily developed for the assessment of systemic effects after oral exposure. However, its applicability to other toxicity endpoints and routes of exposure has been evaluated and suggested. For example, availability of chemicals through dermal exposure is being explored in the context of the applicability of TTC to cosmetic ingredients (Williams et al., 2016). The outcome is expected to provide guidance to extrapolate data from the oral to dermal route by a tiered workflow, which will take into account bioavailability via the different routes i.e. differences in uptake and metabolism.

Internal exposure TTC values, referring to a TTC value derived by considering the fraction of a chemical absorbed and bioavailable, were suggested by Partosch et al. (Partosch et al., 2014)

for the extrapolation from external dose, mainly through the dermal route. In cases where the regulation allows for data waiving due to negligible absorption, with low internal exposure, the internal TTC values would thus provide a basis on which to justify the omitting of data.

The application of the TTC for local effects has been focussed on the dermal sensitisation threshold (DST). This applies the same concept as the TTC, assuming no skin sensitisation is expected if human exposure through dermal contact is sufficiently low. It was developed and further refined by Safford et al. (Safford, Api, Roberts, & Lalko, 2015; Safford, Aptula, & Gilmour, 2011; Safford, 2008) through the analysis of the skin sensitising potencies of chemicals, expressed as EC3 values, which is the effective concentration of a test chemical stimulating the proliferation of lymph node cell threefold in the Local Lymph Node Assay (LLNA). The Scientific Committee on Consumer Products (SCCP), in the opinion on the applicability of TTC for the safety assessment of cosmetic ingredients (EC, 2008), concluded that the DST could be potentially applicable for the induction of sensitisation, but not for elicitation of sensitisation. In a recent publication, Roberts et al. (Roberts, Api, Safford, & Lalko, 2015) developed an approach to identify chemicals to which the DST should not be applied.

Based on the analysis of a database for local and systemic effects through inhalation exposure, Carthew et al. (Carthew, Clapp, & Gutsell, 2009) suggested respiratory thresholds to characterise aerosol ingredients from consumer products. In addition, a list of chemical structures to which the respiratory TTC should not be applied was identified, among them strong acids and bases (potentially strong irritants). Escher et al. (Escher et al., 2010) have proposed lower TTC values than Carthew et al. (Carthew et al., 2009) for systemic effects after inhalation exposure, based on a different database of inhalational toxicity studies. In a recent paper by Schüürmann et al. (Schüürmann, Ebert, Tluczkiewicz, Escher, & Kühne, 2016), structural alerts for repeat dose inhalation toxicity were identified, thus enabling discrimination between high and low potency toxicants through a mechanistic interpretation. The authors suggested that this knowledge could be applied further in risk assessment to derive TTC values tailored to inhalation toxicants.

The use of TTC is being explored extensively and is also gaining support for its application as a screening tool for mixtures risk assessment (Boobis et al., 2011; Koster et al., 2015; Leeman, Krul, & Houben, 2013; Price, Hollnagel, & Zabik, 2009).

Further, there has been an attempt to develop TTC for application in the environmental risk assessment, an ecological threshold of toxicological concern – the so-called ecoTTC. Exposure thresholds to different chemical classes acting through different modes of action (MoA), found in the aquatic compartment, were derived by several authors de Wolf et al. (de Wolf et al., 2005), Tolls et al. 2009 (Tolls, Müller, Willing, & Steber, 2009), Damme et al. (2011) (Damme et al., 2011), (Gross et al., 2009) and Williams et al. (Williams, Berninger, & Brooks, 2011). Gutsell at al. (Gutsell, Hodges, Marshall, & Roberts, 2015) introduced the ecoTTC regardless of the MoA and their analysis supports the use of the ecoTTC for screening purposes. In a recent publication by Belanger et al. (Belanger et al., 2015), under a joint initiative of the International Life Sciences Institute - Health and Environmental Sciences Institute (ILSI-HESI) to develop an ecoTTC, a theoretical overview of the boundaries and challenges was explored. The aim of the latter initiative was to develop an ecoTTC approach that will be accepted for regulatory means.

The Cramer classification has also been tested several times to assess and determine the accuracy of the classification of chemicals, work has been on-going from Patlewicz et al. (Patlewicz, Jeliazkova, Safford, Worth, & Aleksiev, 2008), Lapenna & Worth (Lapenna & Worth, 2011), and Pinalli et al. (Pinalli, Croera, Theobald, & Feigenbaum, 2011). Extensive research has considered the Cramer classification for fragrance ingredients in order to provide guidance and suggest refinements for the revision of the Cramer scheme (Bhatia et al., 2015; Roberts, Aptula, et al., 2015; Schnabel & Taylor, 2015)

Zarn et al. (Zarn, Hänggi, & Engeli, 2015) evaluated non-chemical parameters affecting the TTC values. They studied toxicity data from publicly available evaluation reports for plant protection products, from European Food Safety Authority (EFSA), Joint Meeting on Pesticide Residues (JMPR) of Food and Agriculture Organization (FAO)/World Health Organization

(WHO), and the United States Environmental Protection Agency (US EPA), to assess the impact of study design and NOEL distributions to TTC values. The results of the analysis demonstrate that the lowest 5th percentile NOEL is influenced by the study design and in particular by the dose spacing in the study design and the availability of toxicological studies per chemical.

For regulatory purposes, the TTC approach has been so far applied successfully in the safety assessment of food contaminants migrating from packaging by the United States Food and Drug Administration (US FDA), as well as flavouring agents by the Joint FAO/WHO Expert Committee on Food Additives (JECFA). EFSA uses the TTC approach to evaluate flavouring substances, and the European Medicines Agency (EMA) uses it in support of evaluations of genotoxic impurities in pharmaceutical preparations. The TTC approach is also applicable in the assessment of ground water contaminants, including pesticide metabolites. Therefore, from a relative conservative starting point, the concept has been applied in a number of risk assessment scenarios and has been applied by a number of agencies and regulatory bodies.

1.3 European Union Legislative Frameworks: PPP, Biocides, Cosmetics

Some naturally occurring and man-made chemicals (xenobiotics), foreign to biological systems, have adverse effects on human health. With the purpose to protect individual's health from exposures to such chemicals, several pieces of regulation have been put in place. Those regulations require the careful evaluation of the safety of a chemical before it can be included on the list of authorised substances to be used further in products.

Regulations concerning the placing on the market of plant protection products (PPP), biocidal products and cosmetics are relevant for the present study, therefore they are discussed further within this context in the following sections.

1.3.1 The Plant Protection Products Regulation

Regulation (EU) 1107/2009 (EU, 2009) sets out the requirements, procedure and timeframes for authorisation of Plant Protection Products (PPPs).

Following the requirements under the *European Parliament and European Council Regulation for placing Plant Protection Products on the Market ((EC) No 1107/2009)*, active ingredients of plant protection products need to undergo careful evaluation of safety before they can be included on the list of authorised substances to be used in plant protection products. Approved active substances are included on the list maintained by the European Commission and made available to the public through a web interface *(http://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/public/?event=homepage&language=EN).*

For the evaluation of safety, a range of toxicological studies is required for active substance, and a safener and a synergist if they are added as co-formulants to the plant protection product. In such cases, where specific regulatory requirements exist, the TTC approach cannot be applied. However, as the use of the TTC approach is supported for the assessment of the relevance of metabolites and degradates of pesticide active ingredients for dietary risk assessment, data from toxicological studies on plant protection products active ingredients can be applied for the evaluation of the applicability of *in silico* alternatives. This issue is further discussed and investigated in Chapters 2 and 3 of this thesis.

1.3.2 The Biocidal Products Regulation

Biocidal products can be placed on the market according to the *Biocidal Products Regulation (BPR, Regulation (EU) 528/2012)*. This Regulation has been applicable from 1st September 2013 repealing the *Biocidal Products Directive (Directive 98/8/EC)*. The BPR harmonises the market at the European Union level as the products can be also authorised at the European Union level. It simplifies the authorisation of active substances and introduces timelines for Member State evaluations, opinion forming and decision-making. It also ensures high levels of protection of human health (EU, 2012). As regards data requirements, it is possible to waive requirements if data are not scientifically necessary, if they are technically impossible to supply, or if they are not relevant (exposure associated with the proposed uses). For regulatory purposes 'exposure based waiving' (EBW) means exemption from conducting studies on hazardous properties of chemicals, when the justification to avoid them is based on the fact that there is no relevant

exposure of humans and environment expected. Relevant exposure means that exposure remains within acceptable burden limits, and that it can be assumed the exposure is not associated with any hazard potential for human health and the environment (ECHA, 2010). In accordance with the Article 6 of the BPR, an applicant does not need to provide data as part of the dossier for the approval of an active substance, if the data are not necessary owing to the exposure associated with the proposed uses. In particular, when there is limited information on repeated dose toxicity and/or reproduction, in light of the desire to reduce testing on vertebrates, regulation indicates the use of non-testing methods and provides the opportunity to waive testing, based on exposure considerations. Threshold of Toxicological Concern (TTC) is an alternative method that could give an estimate if EBW can be applied.

According to the BPR (EU, 2012), the biocidal products are defined as follows and are intended to exclude plant protection products, namely pesticides:

1. Any substance or mixture, in the form in which it is supplied to the user, consisting of, containing or generating one or more active substances, with the intention of destroying, deterring, rendering harmless, preventing the action of, or otherwise exerting a controlling effect on, any harmful organism by any means other than mere physical or mechanical action.

2. Any substance or mixture, generated from substances or mixtures which do not themselves fall under the first definition, to be used with the intention of destroying, deterring, rendering harmless, preventing the action of, or otherwise exerting a controlling effect on, any harmful organism by any means other than mere physical or mechanical action.

1.3.3 Regulation on Cosmetic Products

In the European Union, the *Regulation (EC) No 1223/2009 on Cosmetic Products* was passed to ensure that cosmetic products placed on the market are safe for the consumer. In order to assure a maximum level of consumer protection, the regulation requires from manufacturers

preparation of product safety reports, where all potential hazards are carefully addressed, evaluated, and considered in terms of reasonable use of the product. Due to hazardous properties of substances classified for carcinogenicity, mutagenicity and reproductive toxicity category 1 and 2, the regulation prohibits their use in cosmetic products. The regulation also requires that all the products that are to be placed on the European market have to be first registered with the *European Commission* through the *Cosmetic Products Notification Portal (CPNP)*. The information for each product is then made available to competent authorities and poison centres in European member states, which allows constant market surveillance across the European Union.

The regulation (*Regulation (EC) No 1223/2009)* bans animal testing of products for cosmetic purposes. A ban on animal testing of cosmetic ingredients means that the hazard properties of cosmetic ingredients have to be inferred through alternative methods to animal testing, that is through *in vitro* or *in silico* methods. The first provisions on the ban on sale of cosmetic ingredients and their formulations tested on animals were laid down in the *6th Amendment* to the *Cosmetic Directive 76/768/EEC* in 1993 which was delayed up until the *7th Amendment* in 2003 which contained a phased-in ban on animal testing for cosmetics. There are numerous initiatives such as research activities under the "Safety Evaluation Ultimately Replacing Animal Testing" (SEURAT-1) Cluster of projects (Gocht & Schwarz, 2014) and the European Partnership for Alternative Approaches to Animal Testing (EPAA), aimed to develop, validate and make acceptable alternative approaches to animal use in regulatory testing (Cozigou et al., 2015) An international cooperation of validation bodies was set up, the "International Cooperation on Cosmetics Regulation" (ICCR). The initiatives were developed in parallel to the legislative requirements, with the aim to develop scientifically sound and valid alternative methods for hazard assessment that would be acceptable for regulatory purposes. Ten years after the introduction of the ban on animal testing, in March 2013, a full ban on testing for cosmetic purposes entered into force, regardless of the availability of alternative testing methods.

One of the most recent projects, the COSMOS project (2011-2015), was dedicated to the integration of *in silico* models for the prediction of human repeated dose toxicity. For this research initiative, the *European Commission* and the cosmetic industry have joined efforts to attempt to fill current gaps in scientific knowledge and accelerate the development of non-animal test methods, particularly relating to the complex human health related endpoints such as repeat dose toxicity. One key aspect of COSMOS activities included the extension of the concept of Threshold of Toxicological Concern (TTC) to assess cosmetic ingredients (Yang, Cheeseman, & Worth, 2014).

1.4 Computational Methods to Predict Toxicity

Hazard properties of chemicals are traditionally inferred from *in vivo* toxicological testing. *In vivo* testing has a long history and it is thought as being the most representative for human toxicity. However, especially with regard to *in vitro* testing, alternative methods, which are potentially acceptable for regulatory purposes, also exist.

If toxicological data from toxicity testing cannot be obtained, or toxicological data required by regulation could be waived due to negligible exposure, (see discussion above regarding the Cosmetics Regulation) then, in such cases, *in silico* methods may provide a viable alternative. *In silico* approaches include computer-based models to predict a broad range of toxicity endpoints, genotoxicity among them ((Serafimova, Gatnik, & Worth, 2010), (Enoch & Cronin, 2010)).

Computational methods are a fast evolving area under the replacement of animal testing strategy. They offer a rational and consistent way to predict, based on the chemical structure alone, physico-chemical properties, toxicological endpoints and other biological effects, as well as fate in the environment and biological organisms. In order to facilitate their use, (Q)SAR models have been integrated in software tools and made accessible through user-friendly interfaces. A wide range of software tools is available commercially and as freeware (Fuart Gatnik & Worth, 2010). As it is difficult to infer human relevance directly from a single alternative approach *in silico* approaches should be seen as part of Weight of Evidence approach

(WoE) from which conclusions can be used within an Integrated Testing Strategy (ITS) (Ellison, Madden, Judson, & Cronin, 2010).

The Threshold of Toxicological Concern (TTC) approach, first formalised for risk assessment by Kroes (Kroes et al., 2004) relies on the stepwise assessment of genotoxicity (and hence genotoxic carcinogenicity), the anti-acetylcholinesterase effects of organophosphates and carbamates, as well non-cancer structural hazard classes, Cramer classes. Therefore (Q)SAR models to predict genotoxicity, as well as the computational implementation of the Cramer decision tree for classification of chemicals for their expected hazard have been identified to play an important role in the successful application of TTC. To date, several (Q)SAR models have been published in the literature to predict genotoxicity and extensive reviews and evaluation studies are provided in publications from Serafimova et al. (Serafimova et al., 2010), Ellison et al. (Ellison et al., 2011), Fioravanzo et al. (Fioravanzo, Bassan, Pavan, Mostrag-Szlichtyng, & Worth, 2012), Worth et al. 2013, Benigni et al. (Benigni, 2014) Cassano et al. (Cassano et al., 2014).

Other computational methods, such as those to predict physico-chemical properties, and identify chemical substructures present in the datasets have been applied extensively to compare chemical datasets in the sense of exploring the chemical space. Definition of chemical space as described by Worth (2012), is a representation of the structural features and/or molecular properties covered by a defined set of chemicals. The molecular properties may include intrinsic properties (defined purely by chemical structure), such as size and shape, derived properties such as chemical reactivity, as well as extrinsic and biologically relevant properties such as metabolic activity.

By using chemoinformatics methods, it is possible to visualise and characterise chemical space in a consistent manner, so that different datasets (including regulatory inventories and datasets suitable for model development) can be compared. Such comparisons enable regions of overlap and divergence to be identified, as the basis for targeted model development, testing, and/or regulatory action.

It should be noted that the development and application of chemoinformatics methods is an active area of research, and yet there is no single agreed approach for the use of chemical space analysis in toxicology, therefore a number of approaches were considered in this study.

Reviews of the methods and software available for the calculations of physico-chemical properties have been undertaken by several authors: Mostrag-Szlichtyng & Worth (Mostrag-szlichtyng & Worth, 2010), Madden (Madden, 2010), Adler et al. (Adler et al., 2011) and Liao et al. (Liao, Sitzmann, Pugliese, & Nicklaus, 2011). Computational methods applied in the context of this thesis are discussed in detail within the relevant chapters.

1.5 Research Aims of this Thesis

The Systems Toxicology Group at the Joint Research Centre (JRC) provides support in the area of the application of computational methods for toxicity assessment to several bodies and committees of the European Union, such as: EFSA, the Directorate General for Health and Food Safety (DG SANTE), the Scientific Committee on Consumer Safety (SCCS), the Scientific Committee on Health and Environmental Risks (SCHER), and the Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR). In accordance with the research objectives of the group, the overall aim of the thesis was to assess the applicability of the Threshold of Toxicological Concern approach to developmental and reproductive toxicants, cosmetic ingredients and biocides active substances. The research also aimed at proposing a workflow to assess genotoxicity by applying *in silico* methods for the purpose of applying a TTC decision tree. The following is a more detailed description of specific research questions and the rationale behind them. The specific aims of the thesis were:

- *To demonstrate the possible contribution of QSAR models to identify genotoxic nitrobenzenes. The evaluation attempted to identify and suggest the best possible combination of QSAR tools for the prediction of genotoxicty. Chemical space of active and inactive genotoxicants was characterised to further support the evaluation of genotoxicty in a weight of evidence manner.*

The National Institute of Environmental Health Sciences (NIEHS) Integrated Prediction System Alpha Test was used as source of Ames genotoxicity data for nitrobenzenes. Structural files containing 253 nitrobenzenes with Ames data were generated. Seven QSAR software models were applied and evaluated for the prediction of genotoxicity in order to support the application of the TTC exposure thresholds. The findings of this analysis are reported in Chapter 2.

- *To develop a case study with a nitrobenzene pesticide active substance, which is a negative for Ames genotoxicity but predicted as positive by all the QSAR models utilised. The stepwise approach that was applied included a combination of QSAR models to predict genotoxicity. It further suggested the application of category formation and applied read-across to predict genotoxicity.*

Ames test results were gathered from publicly available toxicity databases. A combination of QSAR tools was applied. Then, the formation of deoxyribonucleic acid (DNA) adducts due the formation of nitrenium ions from nitrobenzenes as identified by the profilers within the OECD QSAR Toolbox as crucial in nitrobenezenes' genotoxicity was assessed. Based on that finding a category of similar chemicals was developed. By applying read across, genotoxicity was predicted. The findings of the analysis are reported in Chapter 3

- *To evaluate the application of the TTC approach for the assessment of cosmetic ingredients. The chemical space for cosmetic ingredients was compared with the reference Munro dataset. The Toxtree Cramer classification was applied, followed by the cumulative distribution analysis of the repeat dose NOEL values for cosmetic ingredients.*

A list of cosmetic ingredients with toxicological values and their related structural files were made available to JRC. Some computational tools were employed to describe the chemical space in terms of substructure representations and physico-chemical parameters. Toxicological data were translated into chronic NOEL values in cases where sub-chronic or Lowest Observed Effect level (LOEL) values were available. Subsequently, the TTC structural file was processed

through Toxtree implemented Cramer classifications followed by distribution analysis of toxicological values (NOEL) to derive the 5th percentile NOEL for threshold estimation of each Cramer class. The findings of the analysis are reported in Chapter 4

- *To evaluate the applicability of the TTC approach to biocides by analysing the chemical space for biocides active compounds, Toxtree Cramer classification and the cumulative distribution analysis of biocides active compounds.*

A list of biocides active substances was made available based on which structural files were generated. Then a selection of computational tools was employed to describe the chemical space in terms of substructure representations and physico-chemical parameters. Toxicological data were retrieved from publicly available databases and translated into chronic NOEL values. Following this, the TTC structural file was processed through Toxtree implemented Cramer classification followed by distribution analysis of toxicological values (NOEL) to derive the 5th percentile NOEL for threshold estimation for each Cramer class. The findings of the analysis are reported in Chapter 5

- *To classify compounds into Cramer Classes using Toxtree and analyse the cumulative distribution of developmental toxicants and substances with adverse effects on sexual function and fertility in order to derive threshold values for the specific class of toxicants.*

A list of classified reproductive and developmental toxicants with toxicity data (Muller et al., 2012) was made available by Dutch National Institute for Public Health and the Environment (*RIVM*) through the EFSA TTC working group The chemical structures were not available; therefore, several publicly available sources (Chemspider (http://www.chemspider.com/), Chem ID*plus* (http://chem.sis.nlm.nih.gov/chemidplus/) were used to collate SMILES (Simplified Molecular Input Line Entry System) codes for chemical structures in order to allow further application of the Cramer decision tree implemented in the ToxTree software to assign hazard classes. The cumulative distribution analysis was performed to derive the 5th percentile

NOEL for the specific groups of the classified toxicants. Upon this, threshold values were derived. The findings of the analysis are reported in Chapter 6

2 Application of QSAR Models to Predict the Genotoxicity of Nitrobenzenes

2.1 Introduction

Nitroaromatic compounds are among the largest and most important group of industrial chemicals in use today. The class comprises explosives trinitrotoluene (TNT), pesticides (dinoseb, carbofuran), pharmaceuticals (chloramphenicol), dyes (picric acid - a yellow dye for fabrics), solvents and many other industrial and consumer products, which are produced using nitroaromatic compounds as starting materials (Ju & Parales, 2010). They are reactive chemicals and well known to cause specific adverse health effects; in addition, they are recognised as important environmental pollutants. These compounds may be present in the environment as active ingredients or as impurities at trace levels; therefore several pieces of legislation, aiming to regulate the use and exposure to harmful chemicals, have been put in place to control them.

With regard to this thesis, one such piece of legislation is *Directive 2001/83/EC on the Community Code Relating to Medicinal Products for Human Use*. This Directive lays provisions to assess the safety through toxicity tests, also including mutagenicity, for any new medicinal compound. Subsequently, the EMA guidance document describes the framework and approaches if genotoxic impurities are found in a new active substance. It emphasises that a genotoxic impurity should be identified through the presence of a structural alert [or sub-structural molecular fragment associated with genotoxicity, refer to paragraph on the development of structural alerts (SA) for genotoxicity below], and that it has to be controlled using the Threshold of Toxicological Concern (TTC) approach (EMEA, 2006). Other relevant pieces of legislation include *Regulation (EC) No 1107/2009 Concerning the Placing of Plant Protection Products (PPP) on the Market*; in particular this Regulation lays provisions to assess metabolites arising from the application of a PPP. These metabolites may also be assessed by applying the principle of TTC.

The TTC approach, used for chemicals without toxicological data, builds on generic threshold limits for human exposure. It assumes that the possibility of an adverse effect, the nature of

which cannot be assessed, on human health is negligible below the threshold values. More on the TTC approach can be found in the Section 1.2 of the introductory chapter

Carcinogenicity is assumed the most sensitive endpoint considered, if chemicals are carcinogenic. Therefore thresholds protecting from carcinogenic effects should protect also from other adverse effects (Boobis, 2015). The most potent carcinogens are considered those that act through a genotoxic mode of action, therefore lower threshold values have been identified for this group of chemicals (Cheeseman et al., 1999; Kroes et al., 2004). As genotoxicity can be predicted efficiently from the structure alone, *in silico* methods provide a viable alternative when toxicological data are missing (Enoch & Cronin, 2012). The need to identify a structural alert, which is predictive for DNA reactivity, (Benigni & Bossa, 2011) that raises concern for carcinogenicity, is one of the primary requirements in the decision tree process for the application of the TTC principle (Kroes et al., 2004).

The basis and derivation of the structural alerts for genotoxicity can be considered to be part of the development of *in silico* approaches. An alternative method, *in silico* assessment, which provides a means to assess genotoxicity when toxicological data are lacking, relies on computational models interpreting and applying mechanistically derived chemistry. For example, Structure-Activity Relationships (SARs) and Quantitative Structure Activity Relationships (QSARs), collectively referred to as (Q)SARs, are theoretical models that relate the structure of chemicals to their biological activities. (Q)SARs are developed on the premise that the properties of the chemical depend on its intrinsic nature and can be predicted directly from its molecular structure and inferred from the properties of similar compounds whose activities are known (Worth et al., 2007). (Q)SARs are used to predict the physico-chemical, biological (e.g., toxicological) and fate properties of molecules from a knowledge of chemical structure. In addition to the formalised approach of QSAR analysis, it is possible to estimate chemical properties and endpoints using a less formalised approach based on the grouping and comparison of chemicals (by read-across).

One of the simplest and best known approaches to predict genotoxicity for structurally diverse chemicals is based on the use of structural alerts (SAs), in order to identify electrophilic functional groups or substructures (Serafimova et al., 2010). One of the first set of SAs for carcinogenicity was defined by Ashby and Tennant (Ashby & Tennant, 1991). These definitions followed the electrophilicity theory originally developed by James and Elizabeth Miller who stated: "*that most, if not all, chemical carcinogens either are, or are converted in vivo to, reactive electrophilic derivatives which combine with nucleophilic groups in crucial tissue components, such as nucleic acids and proteins*'' (Miller & Miller, 1981). Since the attack and modification of DNA is the main step in the mechanism of action of many carcinogens (i.e., the so-called genotoxic carcinogens), the SAs related to such classes of carcinogens are also valid for the *Salmonella* mutagenicity (Benigni & Bossa, 2011). Many SAs have been developed for genotoxic mutagenicity, primarily based on considerable Ames test results and interpreted according to electrophilic chemicals; examples of SAs include also nitro-aromatic groups (Enoch & Cronin, 2010) which are associated with well-defined mechanisms of action.

Genotoxicity refers to the process that alters the structure, information content or segregation of DNA. It includes, but is not limited to, mutagenicity, which refers to the induction of permanent transmissible changes in the amount or structure of the genetic material in cells or organisms. Thus, genotoxicity testing is performed to assess the potential of a substance to induce genotoxic effects, which may cause heritable damage or lead to cancer in humans.

Standard genotoxicity tests requested for regulatory purposes (as for example *Regulation (EC) No 1107/2009*) include an *in vitro* test for bacterial reverse mutation (the Ames test), cytogenetic evaluation of chromosomal damage with mammalian cells (chromosomal aberrations) or an *in vitro* mammalian cell gene mutation test (mouse lymphoma thymidin kinase assay). *In vivo* tests for chromosomal damage using rodent hematopoietic cells (*in vivo* chromosomal aberrations and *in vivo* micronucleus assay) are also used.

The analysis presented here is based on Ames test results. The Ames - bacterial reverse mutation - test, uses amino acid requiring strains of *Salmonella typhimurium* and *Escherichia coli* to

detect point mutations (OECD 471, 1997). In this test the mutations in the test strain are reversed to normal functionality that is to the ability to synthesise an essential amino acid histidine.

Extensive studies on genotoxic chemicals provide strong evidence that damage to DNA can occur at very low doses, without an apparent threshold, and that damage increases steadily with increasing dose (COC, 2014). As a consequence, compounds with a structural alert for genotoxicity should be considered separately in the TTC application scheme as the assumed threshold is lower than for non-genotoxic compounds. In the scenarios where a compound is not a member of the exclusion group - the COC and other structural alerts for genotoxicity have been identified, a threshold exposure of 0.15 µg/day can be applied. A further extension/refinement of the TTC decision tree was proposed by Felter et al (Felter et al., 2009), which aimed to also include Ames test data and the duration of exposure (12 months), and concluded that higher exposure levels could be acceptable. The extended part of the decision tree is presented in Section 1.2. of the Introductory chapter. Added steps are represented with dashed borders.

2.1.1 Genotoxicity of Nitroaromatic Compounds

The genotoxic activity of the nitroaromatic compounds is presumed to arise from metabolic activation of the aromatic nitro moiety, probably giving rise to reactive DNA adducts. The metabolism is mediated by P450 enzymes, xanthine oxidase, aldehyde oxidase and quinine reductase. The reductive products may involve the nitro anion radical, nitroso intermediate, hydroxylamine to give rise to corresponding aniline metabolite, most of which are the potential contributors to the toxicity of nitroaromatics. These species have been shown to cause DNA damage in the form of DNA strand breaks. DNA damage by covalent binding is caused by the adduct arising from the activated metabolite *N*-hydroxylamine. As shown in Figure 2.1, the mechanism involves formation of the extremely electrophilic nitrenium ion and subsequent binding with DNA through the nucleophilic substitution - S_N1 mechanism (Enoch & Cronin, 2010; Kalgutkar et al., 2005).

Aromatic hydroxylamine is electrophilic and thus is a itself toxicophore and is suggested to undergo *O*-acetylation, *O*-sulfatation, or *O*-protonation to form electrophilic intermediates that covalently bind to DNA (Benigni & Bossa, 2006; Kazius, McGuire, & Bursi, 2005). They will also directly react with the DNA by the acid-catalysed reaction (Beland & Poirier, 1989)

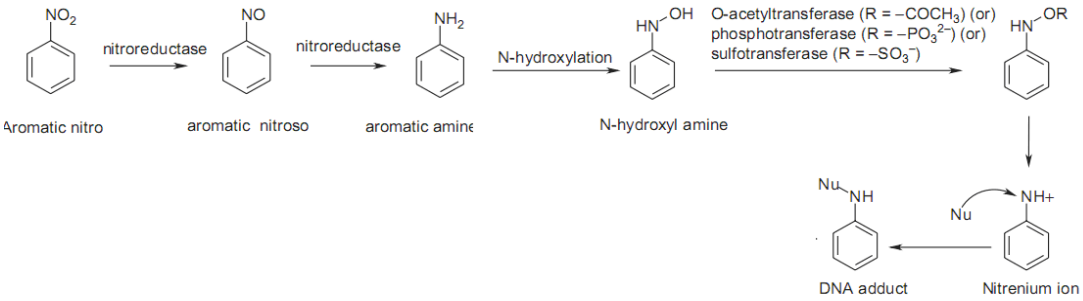


Figure 2.1. Metabolic conversion of aromatic nitro to aniline and subsequent conversion of aniline to electrophilic nitrenium ion, which can then result in a reaction with DNA adopted from Enoch & Cronin, 2010 (Enoch & Cronin, 2010).

2.2 *Aim of the Study*

In silico modelling offers a possibility to predict genotoxicity in the absence of compound specific toxicological data. Nitrobenzenes are compounds also applied as active substances in plant protection products under *Regulation (EC) No 1107/2009.* As the TTC approach is suggested to be applied for the assessment of pesticide metabolites, and includes the identification of genotoxic compounds, the ability of the available (Q)SAR models to predict this specific endpoint was evaluated.

Therefore the aim of this study was to address the outstanding need in respect of developing strategies for the application of alternative methods to identify genotoxic structural alert or predict genotoxicity, thus contributing to the application of the Threshold of Toxicological Concern, as a step in early stages of risk assessment.

2.3 *Data and Methods*

A test set of monocyclic nitroaromatic compounds (termed nitrobenzenes in the text) with experimental Ames data was compiled by extracting the data from National Institute of

Environmental Health Sciences (NIEHS) Integrated Prediction System (IPS) Alpha Test, as was available in August 2011, which is a database with an integrated prediction system. The genotoxicity experimental data are derived from several sources: CPDB, DSSTox, NTP and Chemical Carcinogenesis Research Information System (CCRIS).

2.3.1 Dataset Compilation

The NIEHS Integrated Prediction System Alpha Test was used as source of toxicity data. The NIEHS IPS offers a platform where chemical specific toxicological data and *in silico* tools (methodologies) are available to support the toxicological assessment (http://insilicofirst.com/). The toxicity data in the system originate from several sources - FDA PAFA Database, National Toxicology Program (NTP) Chronic Database, Registry of Toxic Effects on Chemical Substances (RTECS), Distributed Structure-Searchable Toxicity (DSSTox) Database, Carcinogenicity Potency Database (CPDB) and covers acute, multiple dose studies including sub-chronic liver, carcinogenicity, genetic toxicity, reproductive and irritation.

The system requests a login and upon identification, the window with search options is opened up. In this step search only for nitroaromatic structure in the database was performed with substructure search function. A SMILES code for nitrobenzene was inserted in the search window. Then the search button was selected and in the next step, the system offered the possibility to add chemical ID data. Upon loading this additional data, two files were selected for export an .xls and an .sdf file. This first search resulted in two files each containing 5,347 nitroaromatic structures with Chemical Abstracts Service (CAS) registry number and chemical name. The sdf file instead served as a source for structures' SMILES code. The file was opened with the Accelerys discovery studio from where the columns with the SMILES code and CAS were copied into the .xls file of 5,347 structures.

In the second step, no structure was specified in the search window, the search with toxicity endpoint data was used to retrieve structures with available Ames genotoxicity data from the system's database. .xls and .sdf files were selected for export, containing 8,811 chemicals' structures. Also, in this case, the file contained a CAS and a chemical name for each compound.

The two .xls files were matched by CAS number to connect the nitroaromatic structures with Ames genotoxicity data, which resulted in 570 nitroaromatics with experimental *Salmonella* genotoxicity data. Then structures containing more than one nitrobenzene substructure were removed manually. Finally, a dataset of 252 monocyclic nitroaromatic compounds (nitrobenzenes) with Ames test results (174 actives and 78 negatives) was obtained and used to evaluate the predictivity of various QSAR software models for Ames genotoxicity. A .smi file was generated from the .xls file, then Open Babel v 2.2.3. was used to generate the .sdf and .mol structure files for the list of nitrobenzenes. These structural files allowed for the running of the prediction models.

Another compound, sodium-5-nitroguaiacolate, used as a case study developed in Chapter 3, was added to the dataset.

2.3.2 Software Packages

Seven software packages were applied to predict Ames genotoxicity of 253 mono-cyclic nitroaromatic compounds (nitrobenzenes). All tools considered applicable for the purpose of this study are commonly and routinely used by industry and regulatory agencies (Cronin et al., 2003; Fuart Gatnik & Worth, 2010).

The applicability of the genotoxicity models to make predictions for the nitrobenzenes was also assessed according the Cooper statistics and is described in the section below.

The software packages used to predict the genotoxicity of the nitroaromatics included statistical tools: CAESAR (Mutagenicity v 2.0), TOPKAT (v 6.2) and PASS (Professional); hybrid tools: TIMES v2.26.5; and the expert knowledge tools Toxtree (v 2.5.0), Derek Nexus v 2.0 and HazardExpert (Pallas v 3.6.2.1). All tools were run in batch mode.

1. CAESAR, now called VEGA, comprises a series of statistically based models to assess five toxicological endpoints, mutagenicity among them. It was developed and released as an open source software tool as an output of the EU CAESAR project (http://www.caesar-project.eu/) to support the registration of chemicals under REACH

regulation. The model to predict mutagenicity applies a support vector machine approach and is derived on the Kazius/Bursi Ames toxicity database (http://cheminformatics.org/datasets/bursi/). The statistical approach is complemented with the SA as defined in the Benigi Bossa rulebase (Benigni, Bossa, Jeliazkova, Netzeva, & Worth, 2008).

2. TOxicity Prediction by Komputer Assisted Technology - **TOPKAT** is a QSAR-based system, developed by Accelrys Inc. (http://accelrys.com/). The model was developed based on US EPA genetic toxicology data. TOPKAT models are derived using a range of two-dimensional molecular, electronic and spatial descriptors. The QSAR for genotoxicity, a categorical endpoint, was developed by discriminant analysis.

3. The Prediction of Activity Spectra for Substances (PASS) software comprises an embedded SAR model based on the 2D structural formulae of compounds in order to predict many types of biological activity simultaneously. The prediction algorithm applies Bayesian estimates of probabilities for a compound to belong to the classes of active or inactive compounds, respectively (Filimonov & Poroikov, 2008). The predicted activity spectrum is presented in PASS by the list of activities, with probabilities "to be active" *Pa* and "to be inactive" *Pi* calculated for each activity.

4. The OASIS/TIMES software is built on a hybrid approach to predict Ames mutagenicity and chromosomal aberration. It combines expert knowledge to identify SAs followed by statistical methods to consider the remaining part of the molecule. The system also includes a metabolic simulator and gives predictions of mutagenicity with and without metabolic activation.

5. Toxtree is a flexible and user-friendly open-source application that places chemicals into categories and predicts various kinds of toxic effects by applying decision tree approaches. It is freely available from the JRC website: *https://eurl-ecvam.jrc.ec.europa.eu.* The mutagenicity predictions are based on a

decision tree, which implements the Benigni Bossa rulebase (Benigni et al., 2008). It predicts the possibility of carcinogenicity and mutagenicity by discriminant analysis and structural rules.

6. **Deductive Estimation of Risk from Existing Knowledge, Derek for Windows** (DfW) (renamed Derek Nexus) is a SAR-based system developed by Lhasa Ltd, *(https://www.lhasalimited.org/).* It contains structural alerts, covering a wide range of toxicological endpoints in humans, other mammals and bacteria. All the reasoning rules (describe the likelihood of toxicity and DfW prediction) in DfW are based either on hypotheses relating to mechanisms of action of a chemical class or on observed empirical relationships.

7. HazardExpert is a module of the Pallas software developed by CompuDrug *(http://compudrug.com/).* It predicts the toxicity of organic compounds based on toxic fragments. It is a rule-based system with an open knowledge base allowing the user to expand or modify the data on which the toxicity estimation relies.

2.3.3 Evaluation of the Classification Models

The goodness-of-fit of the (Q)SAR classification models, i.e. predicting active or non-active genotoxicants, was evaluated by the Cooper statistics (Cooper, Saracci, & Cole, 1979; Worth & Cronin, 2001). These are summarised in Table 2.1 and include: sensitivity, specificity, concordance (accuracy), positive/negative predictivity. In addition, false positive/false negative rates were calculated from the prediction results. To compare the performance of several classification models, the Receiver Operating Characteristics (ROC) curve is often used by plotting the true positive rate (sensitivity) against the false positives rate (1 - specificity). The characteristics of the curve allow for the easier evaluation of the precision of predictions. The positive and negative predictivity and accuracy were also calculated. Accuracy is calculated from the proportion of the total number of predictions that were correct. As shown in Figure 2.2, a model whose statistical evaluation place it in the top left-hand corner of the plot (e.g. Point C+) is ideal, i.e. it is fully predictive, whereas a model with statistics that place it along the

diagonal (e.g. Point B) is producing predictions not better than chance, a model with below the diagonal line (e.g. Point C) is less predictive than chance.

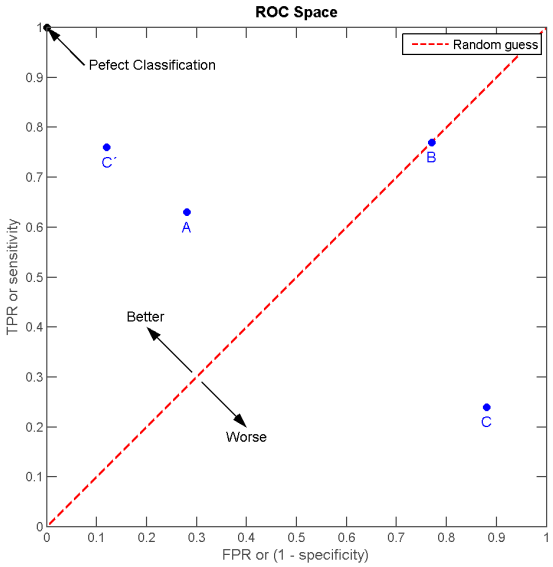


Figure 2.2. Typical Receiver Operating Characteristic (ROC) Space for evaluation of predictive classification models.

Table 2.1. Definition and meaning of statistical terms used to evaluation classification models.

Statistic	Definition	Meaning (proportion/percentage) of …
Sensitivity	TP / (TP+FN)	known positives that are correctly predicted
Specificity	TN / (TN+FP)	known negatives that are correctly predicted
False positive rate = 1-specificity	FP/(TN+FP)	known negatives that are incorrectly predicted as positive
False negative rate = 1-sensitivity	FN/ (TP+FN)	known positives that are incorrectly predicted as negative
Positive predictivity	TP / (TP+FP)	predictions that are true positives (probability of a positive prediction being correct)
Negative predictivity	TN / (TN+FN)	negative predictions that are true negatives (probability of a negative prediction being correct)
Accuracy	TP + TN / (TP+TN+FP+FN)	Total predictions that were correct (probability of a negative and positive prediction to be correct)

2.3.4 Chemical Space Analysis of Genotoxic and Non-Genotoxic Nitrobenzenes

The chemical space of the nitrobenzenes was analysed to assist in the definition of the applicability domains of the models. Chemical space was represented by three physico-chemical properties, reflecting the Lipinski Rule of Fives and describing the size of a molecule – van der

Waals volume (Sv), partitioning (MlogP) and reactivity (the ionisation potential (Mi). The properties were calculated using the DRAGON ver 6 software, a freely available tool developed by Talete Srl. DRAGON 6 calculates a large number of descriptors for structure-activity analysis and those chosen were deemed representative of the important properties governing toxicity for these compounds.

In order to use the software, separate structure (.smi) files, containing a column with SMILES and CAS numbers, were generated. One file contained the 174 active and another the 78 non-active nitrobenzenes. These two files were converted to a 3D .sdf file using Open Babel version 2.2.3 for Windows, freely available at http://openbabel.org. Each file was imported into Dragon separately. Then, the descriptors, namely MLogP, Sv and Mi from amongst the basic descriptors, were selected. The results were stored in a .txt file which was subsequently imported into excel. For the graphical 3D representation of the distribution of calculated descriptors, STATISTICA 10 was used.

2.4 *Results*

A dataset of 174 genotoxic and 79 non-genotoxic nitrobenzenes was constructed that allowed for a robust evaluation of seven QSAR software models predicting genotoxicity for these compounds. Performance statistics were calculated from the predictions and the results are summarised in Table 2.2.

With regard to expert systems for toxicity prediction, the models (i.e. compilations of structural alerts) are likely to have been developed from heterogeneous datasets of expert conclusions based on data from multiple test methods. It is possible and inevitable that the conclusions relating to genotoxic potential will vary between assessors and over time, especially when the criteria for assessing the raw data have changed or were not well defined in the first place. For statistical models, the possible differences in the classification of genotoxicity are a source of variability in the training set, which will inevitably affect the reliability of prediction (Worth, Lapenna, Piparo, & Serafimova, 2011). Without complete reanalysis of the original data, something that is unlikely from a practical point-of-view, it is difficult to eliminate or even

quantify this variability, however, it should be borne in mind when considering performance of models.

Another aspect to consider is the definition and application, or otherwise, of the models' applicability domain. As structural alert models usually lack an adequately defined applicability domain (Ellison et al., 2011), the performance statistics for expert systems and statistical tools were initially evaluated together, without considering whether the compounds were within the applicability domain. For nitrobenzenes, known to have genotoxic potential due the presence of a nitroaromatic moiety, most tools demonstrated high sensitivity, which ranged from 80% to 100%, the last obtained from two expert systems Derek and HazardExpert. It is also important to stress that Derek predicts the probability of genotoxicity based on the presence of a SA and is not able to predict negatives as the validity of the chemical space cannot be assessed. HazardExpert predicted all compounds considered as being genotoxic. A low sensitivity rate, 41%, was only demonstrated for TOPKAT. In the case of expert system tools, high sensitivity is associated with a high false positive rate. Overall, the statistical tools performed better in identifying non-mutagens with specificity ranging from 31 – 71%, the last demonstrated for TOPKAT. Therefore, the results from the expert system tool Derek could be considered as most reliable, although conservative, due the 0% false negative rate. These results could be considered as a strength or advantage, in the context of the application of TTC, when positive (genotoxic) nitrobenzenes should be identified but do run the considerable risk of over-prediction.

The statistical analyses of the *in silico* models were also summarised in a ROC curve (Figure 2.3). Most models are concentrated around the upper right corner, showing high true positive, associated with high false positive, rates of prediction. In an ideal situation, a model with good performance statistics would be placed in the upper left corner on the ROC curve graph, with sensitivity 1 and a 0 false positive rate. Results placing the values for a model on the diagonal imply predictions with the same value as randomly guessing; which is similar to the situation for the predictive capability of TOPKAT.

Table 2.2. Performance of the seven (Q)SAR software models to predict Ames genotoxicity of nitrobenzenes.

	TOPKAT	CAESAR	TIMES	PASS	Derek Nexus	Toxtree	Hazard Expert
ND	2		85		13		
N*	251	253	168	253	240	253	253
TP	66	154	88	173	171	168	174
TN	55	32	18	0	3	12	0
FP	22	47	40	79	66	67	79
FN	94	20	22	0	0	6	0
SE	41	89	80	99	100	97	100
SP	71	41	31	0	4	15	0
FP (1-SP)	29	59	69	100	96	85	100
FN (1-SE)	59	11	20	1	0	3	0
PP	75	77	69	69	72	71	69
NP	37	62	45	0	100	67	0
ACC	51	74	63	68	73	71	69

SE sensitivity, SP specificity, FP false positive rate (1-SP), FN false negative rate (1-SE), PP positive predictivity, NP negative predictivity, ACC accuracy, ND number of compounds not handled by the software, N* number of compounds included in statistics

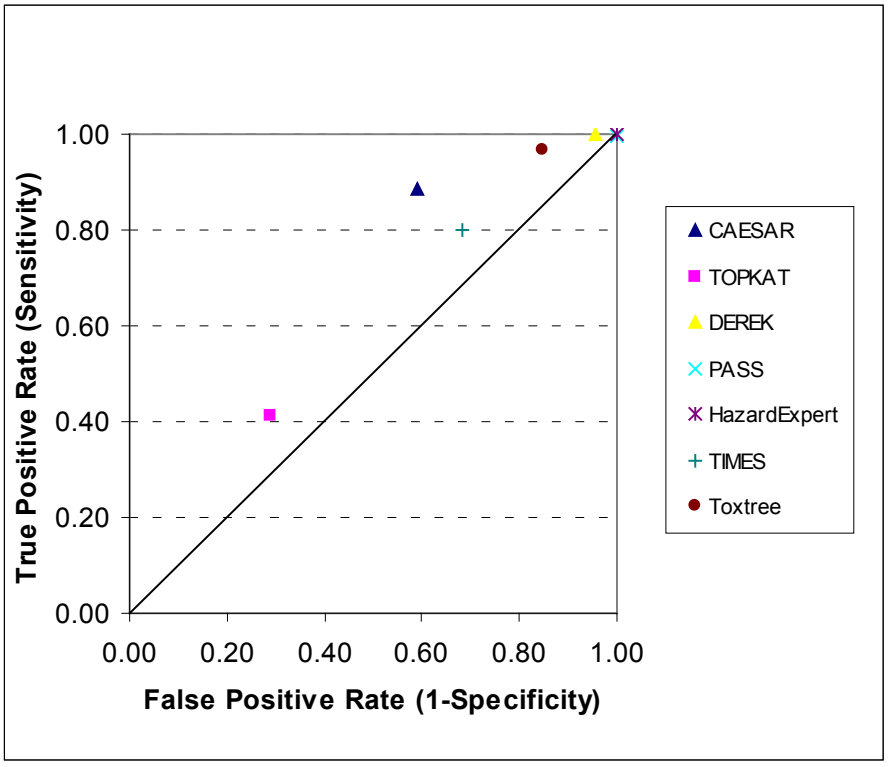


Figure 2.3. ROC curve for monocyclic nitroaromatics Ames genotoxicity prediction from the seven (Q)SAR programs

The applicability domain (AD) of a model is the physico-chemico, structure descriptor and mechanistic space of a model (Netzeva et al., 2005). The defined domain of applicability is where, in theory, the software will make a prediction with a defined reliability. Therefore, when a model is applied to chemicals within its applicability domain, more confidence in the accuracy of predictions is expected. In order to assess the performance of the QSAR models, for those that gave an indication of the AD, this was considered. The predictions for all compounds outside the respective applicability domain for a specific model were excluded from subsequent analysis. The results of the predictions for compounds within the AD are summarised in Table 2.3. Consideration of the AD meant the predictions from CAESAR were improved to give a sensitivity of 96% and specificity 63% with an overall accuracy of predictions of 88%. CAESAR also demonstrated the lowest false negative rate at 4%. TIMES had an overall accuracy of 76%; the false positive rate was zero confirming good positive predictivity, however it fails to identify negatives successfully, with a false negative rate 24% for compounds inside the AD. It should be kept in mind that TIMES has poor coverage for nitrobenzenes (i.e. a very limited AD for these compounds) as 80% of compounds are outside of the AD. TOPKAT was again the worst performing tool with an overall accuracy of 51%.

Table 2.3. Performance of three QSAR software models to predict Ames genotoxicity by considering the assessment of AD

	TOPKAT - all compounds	TOPKAT - within OPS*	CAESAR - all compounds	CAESAR - within AD	TIMES - all compounds	TIMES - within AD
ND	2	2			85	85
N* N**	251	61(24%) 190	253	49 (19%) 204	168	134 (79%) 34
TP	66	60	154	150	88	25
TN	55	36	32	30	18	1
FP	22	21	47	18	40	0
FN	94	73	20	6	22	8
SE	41	45	89	96	80	76
SP	71	63	41	63	31	100
FP (1-SP)	29	37	59	38	69	0
FN (1-SE)	59	55	11	4	20	24
PP	75	74	77	89	69	100
NP	37	33	62	83	45	11
ACC	51	51	74	88	63	76

SE sensitivity, SP specificity, FP false positive rate (1-SP), FN false negative rate (1-SE), PP positive predictivity, NP negative predictivity, ACC accuracy, ND number of compounds not handled or not calculated by the software, N* number of compounds outside AD or Optimum Prediction Space (OPS) limits, Nnumber of compounds included in statistics.**

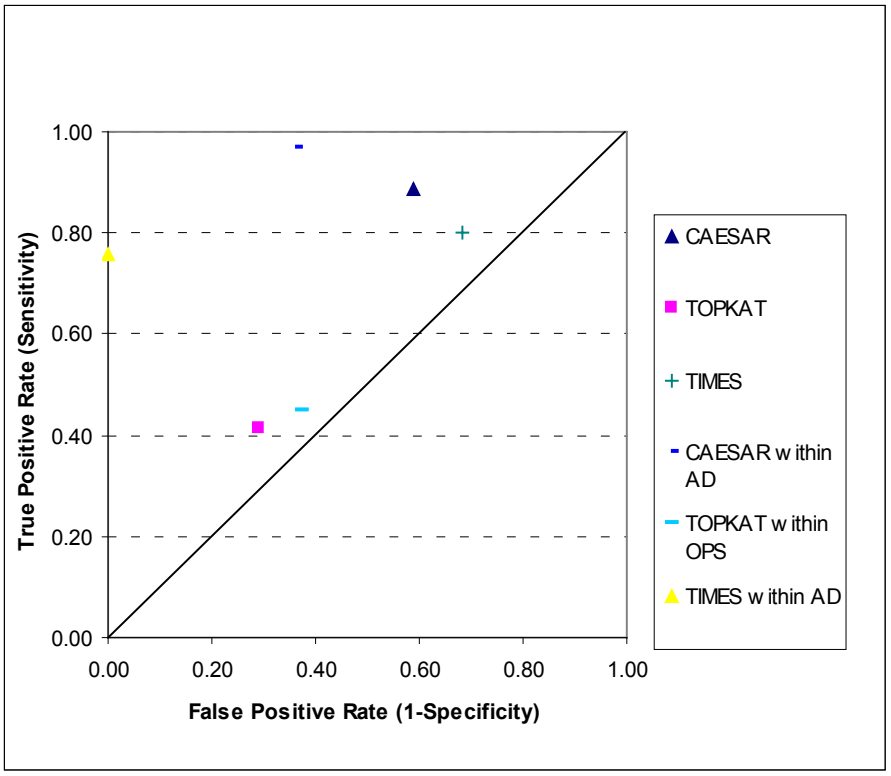


Figure 2.4. ROC curve for monocyclic nitroaromatics Ames genotoxicity prediction following the assessment of AD.

2.4.1 Conclusion: Fitness of (Q)SAR Tools for the Detection of Genotoxic Nitrobenzenes

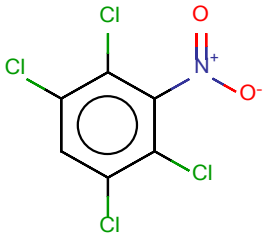
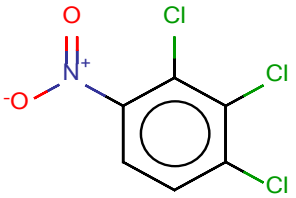
There are many (Q)SAR models to predict genotoxicity, with some of them being assessed for their predictive capability for nitrobenzenes in this study. There is a clear discrimination between statistical and SA based tools with almost all compounds being predicted as positive by the latter. Expert knowledge-derived tools base their predictions on a generally accepted hypothesis, that the nitroaromatic moiety would undergo bioactivation to form a reactive nitrenium ion that produces a DNA adduct (Enoch & Cronin, 2010). In contrast, statistical models have higher negative predictivity, however, it must be borne in mind that these are generalistic models which are not optimised for the nitrobenzenes. Considering the performance

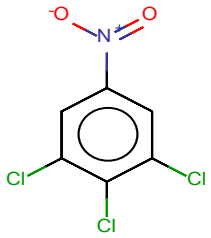
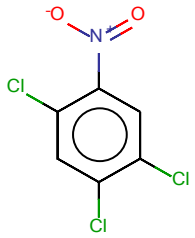
of all predictions (Figure 2.4), CAESAR predicts results for compounds inside the AD with greatest accuracy (88%), followed by Derek, with an accuracy of 73%. Therefore, it could be suggested that a combination of expert systems (Derek) and statistical tools (CAESAR) should be used in an effort to detect genotoxic nitrobenzenes, for applying specific threshold values.

2.4.2 Compounds that were Incorrectly Predicted by all Tools

For compounds listed in Table 2.4 none of the tools was successful in predicting genotoxicity. The compounds belong to the group of polychlorinated nitrobenzenes, which were tested positive for genotoxicity in the Ames test. The compounds presented have a common substituent on the benzene ring, the halogen – chlorine. One possible explanation is that this compounds can probably react directly with DNA through a nucleophilic aromatic substitution (Błaziak, Danikiewicz, & Mąkosza, 2016; Gupta, Saini, & Juneja, 1997). The nitro group would deactivate the ring thus making it susceptible to an attack from a nucleophile. In this case the mechanism related to the reduction of the nitro group would not be the reason for the genotoxicity observed for compounds presented below.

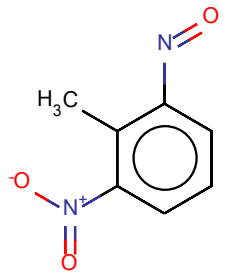
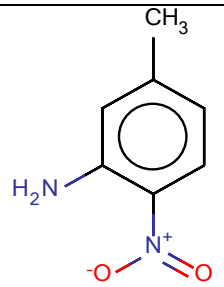
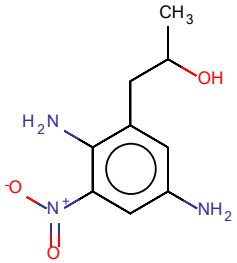
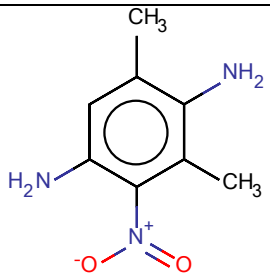
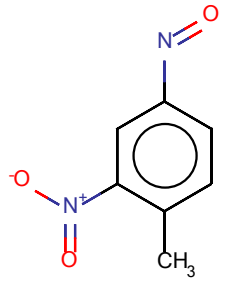
Table 2.4. Four compounds tested as Ames positive, but predicted as negatives.

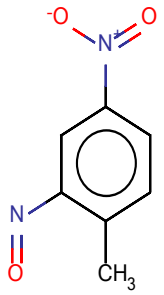
CAS	Chemical name	Structure
117-18-0	1,2,4,5-Tetrachloro-3-nitro-benzene	
17700-09-3	4-Nitro-1,2,3-trichlorobenzene	

CAS	Chemical name	Structure
20098-48-0	1,2,3-Trichloro-5-nitrobenzene	
89-69-0	5-Nitro-1,2,4-trichlorobenzene	

In addition, a number of compounds were found to be inactive in *Salmonella* but were predicted as being positive by all statistical/hybrid tools. The structures of these compounds are shown in Table 2.5. Neumann (Neumann, 2005) discussed that the *in vitro* genotoxicity of monocyclic amino and nitro compounds is not marked. The genotoxic effect alone cannot account for the carcinogenicity of these compounds by itself, therefore more attention should be given to negative *in vitro* genotoxicity results. Neumann further discussed that the bacterial mutagenicity test (standard Ames test), yields negative results, even when metabolic activation promoted by addition of enzymes (S9) is considered, not only for aniline derivatives, *o*-toluidine and 2-nitrotoluene but also for *p*-toluidine and 3-nitrotoluene.. As positive results can be obtained in most cases by the addition of norharman, a comutagen which forms mutagenic compounds from aromatic amines during metabolic activation, Neumann assumed, that the initial substances have weak genotoxic potential which is expressed when the metabolic pathway via the *N*-hydroxylamine and subsequent *O*-acetylation is sufficiently active. This could be a potential explanation for the negative test results reported for the compounds in Table 2.5.

Table 2.5. Non-genotoxic compounds predicted as being active by all software tools

CAS	Chemical name	Structure
143922-95-6	2-Nitroso-6-nitrotoluene	
578-46-1	5-Methyl-2-nitroaniline	
104535-31-1	4-Amino-3-nitro-5-β-hydroxypropylaniline	
97629-64-6	4-Amino-3-nitro-2,6-dimethylaniline	
82414-03-7	4-Nitroso-2-nitrotoluene	

CAS	Chemical name	Structure
82414-02-6	2-Nitroso-4-nitrotoluene	

2.4.3 Chemical Space Analysis

In the chemical space analysis, a selection of physico-chemical properties was analysed to understand better the probable relationship between toxicity and a specific physico-chemical descriptor. Descriptors were chosen based on Lipinski rules with the aim to assess the likelihood of absorption and penetration. It was hypothesised that this might explain, at least to some extent, the activity of selected nitrobenzenes. The estimated logarithm of the octanol-water partition coefficient (MLogP), van der Waals volume of a molecule (Sv) and reactivity in terms of the first mean ionisation potential (Mi) were calculated by using the freely available DRAGON software as a web application available from: http://www.talete.mi.it/products/dragon_description.htm. The results of this analysis, the chemical space distribution of active and non-active nitrobenzenes, are visualised in Figure 2.5. Although there is no obvious separation of actives from non-actives in the chemical space defined by the three properties, it could still be observed that some compounds with a combination of higher lipophilicity (as defined by MlogP) and greater volume were non-genotoxic as defined by Ames. These include some polychlorinated and polyfluorinated nitrobenzenes.

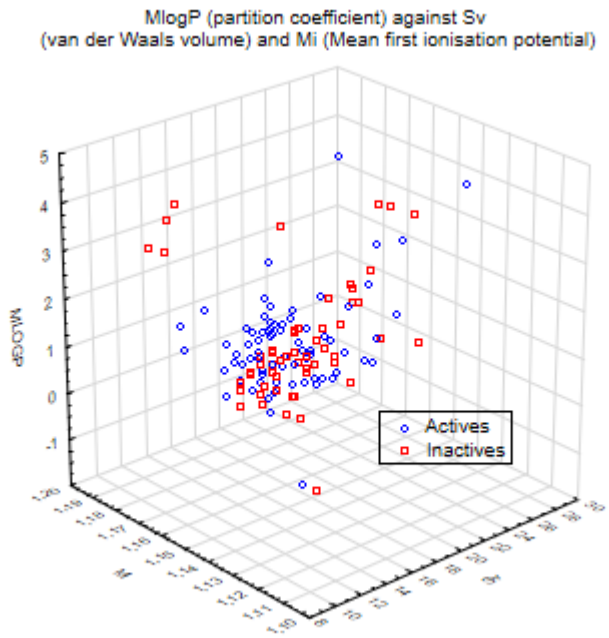


Figure 2.5. Distribution of MLogP, Sv and Mi for genotoxic and non-genotoxic nitrobenzenes

2.5 *Discussion*

For nitrobenzenes, the presumed mode of action associated with cancer in humans is through genotoxicity. The key events leading to the formation of DNA adducts are well understood and implemented in *in silico* prediction models (Enoch & Cronin, 2010). The use of *in silico* prediction methods such as (quantitative) structure-activity relationship models ((Q)SARs), in the context of Threshold of Toxicological Concern (TTC) decision scheme, is intended to be a part of a weight-of-evidence based assessment, to identify specific health concerns, that require the application of specific thresholds. In the scheme proposed by Kroes et al. (Kroes et al., 2004) for the application of TTC for chemicals containing a structural alert for potential genotoxicity, a TTC of 0.15 µg/person per day should be applied following by 1.5 µg/person per day that is also protective for unrecognised carcinogenic compounds. Following the exclusion of highly potent chemicals, other forms of toxicity are considered in the following steps.

Traditional (and regulatory) toxicology relies heavily on the use of *in vivo* data obtained from animal models. (Q)SARs provide a possibility to minimise the reliance on animal testing thus contributing to the efficiency and effectiveness of the risk assessment process. When applying

the TTC scheme, at a certain point, structural alerts for genotoxicity need to be identified. This assumption was considered equivalent to the prediction of genotoxicity by QSAR tools. Therefore, results were assessed in the light of the ability of tools to predict this critical health concern in order to allow the application of lower threshold values.

In this study, the ability of seven (Q)SAR models to predict genotoxicity of nitrobenzenes was assessed. Nitrobenzenes were selected as they constitute an important class of environmental pollutants arising from the application of pesticides and pharmaceuticals among others, where TTC is already applied. Three expert system models, Derek, Hazard Expert and Toxtree, a hybrid tool TIMES, and the CAESAR, TOPKAT and PASS statistical models were assessed for their ability to predict the genotoxicity of nitrobenzenes. The models selected were considered scientifically valid and relevant for regulatory purposes, which allowed a robust evaluation. The statistical analysis of the software performance has shown that CAESAR has the greatest accuracy of predictions for Ames genotoxicity, meaning that in 88% of the cases the prediction for compounds inside the models applicability domain (positive or negative) was correct. Due the critical importance of eliminating false negatives, which are genotoxic compounds predicted to be non-genotoxic, results from Derek should also be considered. In particular, the mechanistic explanation of the results is an asset, as it gives insight into the possible cause of genotoxicity based on which other non-testing methods could be considered, such as use of data on similar compounds. Mechanistic knowledge could also support further expert evaluation. Derek demonstrated a positive predictivity of 72% and a 100% negative predictivity, an overall accuracy of 73%.

As already discussed by Worth et al. (Worth et al., 2011), the crucial question in the application of (Q)SAR is whether any model, or combination of models, is "good enough" for the regulatory purpose (in this case the identification of potential genotoxins). This cannot be answered in the absence of clearly defined performance criteria, and these should be set by the risk assessor and risk manager. For the purpose of risk assessment, where the aim is to diminish the risk, the rate of false negative predictions should be minimised. The analysis presented found that for the

Ames mutagenicity for nitrobenzenes the false negative rate ranges from 0% to 55%. Therefore, a combination of expert system and statistical models, such as CAESAR and Derek, with false negatives rate of 4% and 0% respectively, is suggested. This is precautionary and could efficiently contribute to the risk assessment of nitrobenzenes – i.e. identifying nitrobenzenes with potential genotoxicity in order to apply lower threshold risk dose.

Nitrobenzenes are known to induce cancer through a non-genotoxic mode of action (Hsu et al., 2007). Hsu et al reported indirect and direct effects on the liver, where in the case of direct effects reactive oxygen species are formed or metabolites able to bind to macromolecules. These are non-genotoxic effects and not detectable with the Ames *Salmonella* assay. Therefore, when the *in silico* prediction is that a nitrobenzene is non-genotoxic, there was a second level of protection threshold of 1.5µg/person per day intended to protect from other non-genotoxic carcinogenic effects, which are assumed to occur at higher exposure levels.

However it has to be noted that the recent EFSA, WHO review of the TTC approach (EFSA & WHO, 2016a) supports the removal of 1.5µg/person per day level from the TTC decision tree, as it was concluded that it is of historical value and limited practical value (EFSA, 2012b). The decision, to remove the threshold for non-genotoxic carcinogens is based on the assumption that non-genotoxic carcinogens have a threshold and that doses producing the tumour are in the range of other toxicities, therefore the current Cramer threshold would be protective for non-genotoxic cancer endpoints.

In this chapter the ability of (Q)SAR models to identify SA and/or predict genotoxicity (mutagenicity) was assessed and it was demonstrated that a combination of expert knowledge tools and statistical models give results that can be applied under the WoE approach for the application of TTC. Based on this evaluation a case study, in the following chapter was developed.

3 A Case Study - The Application of QSAR Tools for the Evaluation of the Genotoxicity of a Plant Protection Product, Sodium 5-nitroguaiacolate

3.1 Introduction

Sodium 2-methoxy-5-nitrophenol, or sodium 5-nitroguaiacolate, is a synthetic nitrophenol and is used as an active substance in a plant protection product. It is usually sold in a formulation with sodium *o*-nitrophenolate and sodium *p*-nitrophenolate, under the commercial name Atonik. The product is applied to enhance plant growth and thus increase production yield.

3.1.1 Regulatory Background

Following the requirements under the *European Parliament and of the Council Regulation for placing Plant Protection Products on the Market ((EC) No 1107/2009)*, active ingredients of plant protection products need to undergo careful evaluation of their safety before they can be included on the list of authorised substances to be used in plant protection products. For the approval of an active substance a battery of *in vitro* and *in vivo* tests for genotoxicity, as described in Chapter 2 is required. Usually TTC cannot be applied where specific regulatory requirements exist; however the application of TTC approach has been recommended as a screening tool for the evaluation of Toxicological Relevance of Pesticide Metabolites for Dietary Risk Assessment (EFSA 2012). For this specific purpose, the availability of a full toxicological profile for an active substance, makes such substances good study cases.

Approved active substances are included on the list maintained by the European Commission and made available to the public through a web interface[1]. Active substances are approved for a period no longer than 10 years (EU, 2009)(Regulation (EC) No 1107/2009). After that period an active substance might undergo a renewal procedure. An application for approval is initially submitted to a designated member state – the *Rapporteur Member State (RMS)* that also prepares

[1] http://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/public/?event=activesubstance.selection&language=EN

the *Draft Assessment Report (DAR)*, which is then finally reviewed by the *European Food Safety Authority* (EFSA) based on which EFSA issues the conclusion from the risk assessment evaluation of the active substance. The *Standing Committee on the Food Chain and Animal Health* votes on approval or non-approval of an active substance for specific uses, which is then adopted by the *European Commission*. The procedure is finalised by the publication of the Regulation in the Official Journal of the European Union and the active substance is included on the list of approved substances. After the approval of a substance, a summary of the substance evaluation is made publicly available through an EFSA Scientific report.

3.1.2 Evaluation of Genotoxicity with (Q)SAR, Chemical Grouping and Read-across

As discussed in Chapter 2, alternative approaches to identify genotoxicity or to predict it, for example through the use of structural alerts or other (Q)SARs, exist. Additionally, in this chapter, the assessment by chemical grouping and read-across was also considered.

The main concept behind chemical category formation and read-across is that toxicity of a chemical with unknown toxicological profile is inferred from a group of similar chemicals with known toxicological profiles.

A chemical category is a group of related chemicals with common properties or trends in properties. Members of a category can be related by chemical structure and/or by mode or mechanism of action. The common features allow for the application of read-across to be made between chemicals and endpoints, thereby enabling the filling of data gaps, by applying interpolation or extrapolation of data or more sophisticated structure activity-relationships such as trend analysis and QSARs. The category approach is intended to increase the efficiency of the hazard assessment process (Puzyn, Leszczynski, & Cronin, 2010).

3.1.3 Information on the Compound of Interest

The structure and main physico-chemical properties of the target compound are summarised in the Table 3.1. These were derived from the conclusion on the pesticide peer review for the active

substance (EFSA, 2008): Conclusion on the peer review of sodium 5-nitroguaiacolate, sodium *o*-nitrophenolate and sodium *p*-nitrophenolate.

Table 3.1. Characterisation of the target compound

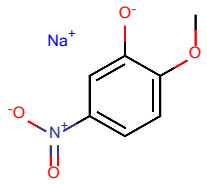
Structural formula	
Chemical name (IUPAC)	2-methoxy-5-nitrobenzen-1-olate sodium salt
Chemical name (CA)	Sodium 5-nitroguaiacolate
CAS No	67223-85-6
SMILES	[Na+].COC1=CC=C(C=C1[O-])[N+]([O-])=O
Molecular formula	$NaC_7H_6NO_4$
Molecular weight	191.12 g/mol

Table 3.2 summarises the results of a standard battery of tests which were performed to detect potential genotoxicity of sodium 5-nitroguaiacolate as a standard regulatory requirement. These were extracted from the EFSA report, cited above.

Based on these results the conclusion of the peer review for sodium 5-nitroguaiacolate is that: "even though some positive results were observed during the *in vitro* genotoxicity studies, the negative results in the *in vivo* testing were supported by the absence of a carcinogenic potential in the long term studies." In this summary, the results regarding the long-term carcinogenicity studies relate to a mixture of all three nitrobenzene compounds, a plant protection product (PPP) containing sodium 5-nitroguaiacolate, sodium *o*-nitrophenolate and sodium *p*-nitrophenolate.

Table 3.2. Experimental data for *in vitro* and *in vivo* genotoxicity, collected by EFSA (EFSA, 2008)

Test system	Test object	Results
In vitro	Ames test - *Salmonella Typhimurium* TA 98, TA 100, TA 1535, TA 1537, TA 1538 *Escherichia Coli*	Negative

	Bacillus subtilis rec assay Strains H17, M45	Negative
	In vitro mammalian cell gene mutation assay	Weak positive in the presence of metabolic activation at the highest concentration
	In vitro chromosome aberration assay	Negative
In vivo	*In vivo* micronucleus assay bone marrow cells of mice	Negative

3.2 Aim of the Study

Sodium 2-methoxy-5-nitrophenol is a pesticide active substance, therefore the findings of a full battery of toxicity tests was submitted at the time of its registration. The availability of toxicological data makes the compound a good case study to evaluate the application of computational tools to predict its potential genotoxicity, thus to support decision making, when applying the TTC decision tree for setting exposure thresholds (Felter et al., 2009; Kroes et al., 2004). This specific example illustrates how difficult it can be to conclude that a chemical compound has a genotoxic potential although a structural alert has been identified.

3.3 Methods

3.3.1 File Generation

Sodium 2-methoxy-5-nitrophenol is a salt. In order to use relevant software for predicting the toxicity of the compound, the SMILES identifier needs to be generated. Additionally a SMILES code was generated for a neutral form of the chemical. Therefore, the sodium was removed and the ionic form was neutralised. The resulting SMILES string was saved as .smi file. The .smi file was then converted into a .mol file using Open Babel. Both types of structure files are accepted by the QSAR tools applied.

3.3.2 Predicting Genotoxicity (Mutagenicity) by Applying Several Software Tools

Three QSAR software tools were applied to predict the genotoxic potential of sodium 5-nitroguaiacolate. Among them, a tool based on expert rules (Derek for Windows v 12.0.0 (now Derek Nexus)), and OncoLogic 7.0) and CAESAR (now named VEGA).

3.3.3 Estimating the Genotoxicity of Sodium 5- nitroguaiacolate by Chemical Grouping

The next step involved the analysis of genotoxicity by chemical grouping and read-across using the freely available OECD QSAR Toolbox (version 2.0.60.597). The OECD Toolbox is a software application used for filling gaps in (eco)toxicity data needed to assess the hazard(s) of chemicals. Crucial to the Toolbox workflow is the grouping of chemicals into chemical categories. The key features include the identification of relevant structural characteristics and mechanism, or mode, of action of a compound of interest. Then compounds with the same structural and mechanistic characteristics are selected and existing experimental data are used to fill data gaps (Cronin & Madden, 2010).

Sodium-5-nitroguaiacolate is presumed to undergo metabolic activation of the nitro group (Enoch & Cronin, 2010) probably giving rise to the electrophilic nitrenium ion, a reactive DNA adduct. This mechanistic knowledge, explained in detail in Chapter 2, will be exploited to build a category.

*3.4 **Results***

3.4.1 Results of Predicted Genotoxicity (Mutagenicity)

All the QSAR tools applied predicted the compound to be active for genotoxicity. Specific results are discussed in detail below (Table 3.3).

Table 3.3. Applied QSAR software tools and prediction results

Software	Prediction	Additional information
DfW (Lhasa Ltd.) http://www.lhasalimited.org	Active	It predicts the compound to be PLAUSIBLE for *in vitro* mutagenicity in bacteria. Alert 329 fired.
OncoLogic™ http://www.epa.gov/oppt/newchems/tools/oncologic.htm	Moderate	Carcinogenic potential assessed as for aromatic amine.
CAESAR http://www.caesar-project.eu/index.php?page=index	Active	Outside Applicability Domain therefore no prediction is provided

Derek for Windows predicted *in vitro* bacterial mutagenicity to be plausible due to the aromatic nitro compound alert (Alert 329). The rationale of the prediction is based on the knowledge that the nitro group undergoes reduction with nitroreductase giving rise to the nitrenium ion, an electrophile that is capable of binding to cellular nucleophiles, including DNA. Nevertheless, it is worth noting that the degree of mutagenicity is dependent upon many factors, including: 1) the rate of reduction of the nitro group, 2) the rate of removal of metabolic intermediates through other biochemical pathways, and 3) the stabilisation of the nitrenium ion by ring substituents. Consequently, a number of substituted nitro compounds are not active in the Ames test (for example ortho/para-nitrophenols (Eichenbaum et al. 2009; Enoch and Cronin 2010)

OncoLogic makes the predictions using a similar rationale to Derek for Windows. It assumes the nitro group will be reduced by nitroreductase to the amino and thus yield an aromatic amine. It further assumes the compound to be an aromatic amine predicting the compound will undergo bioactivation and thus giving rise to electrophilic reactive intermediates, which are capable of interacting with cellular nucleophiles such as DNA. It concludes that the compound, by not having methoxy groups in the ortho position to the amino group, is of moderate concern.

The prediction from **CAESAR** is based initially on a statistical model and then an "expert knowledge" filter is applied based on Benigni/Bossa rules. Therefore the prediction was, as expected, positive. The prediction from CAESAR was however, considered unreliable as the compound is outside its applicability domain. In addition, a non-mutagen from its training set, a 3-nitroanisole, was predicted as a mutagen, therefore showing possible lack of validity of this approach.

In summary, three different software tools were applied to identify a structural alert for genotoxicity and predict the bacterial mutagenicity of sodium 5-nitroguaiacolate; all predicted the compound as being active. The expert knowledge-based tools give the active prediction due to the bioactivation of nitro group. They all predicted the formation of an electrophilic nitrenium ion, capable of forming DNA adducts. It can be concluded that the structural alert for genotoxicity exists in the form of a nitroaromatic moiety.

3.4.2 Predictions of Genotoxicity from Chemical Grouping and Read-Across

The structural alert identified and the background mechanistic knowledge of how nitroaromatic compounds could give rise to DNA adducts was used to evaluate the compound's genotoxicity by chemical grouping methods further. Thus by populating the category formed for sodium-5-nitroguaiacolate with compounds that putatively share the same mechanism of toxicity, accurate predictions should be obtained. The workflow presented bellow was utilised in this study and applied in the Toolbox.

The first step in the OECD QSAR Toolbox workflow is the profiling of the compound. For the purpose of assessing the genotoxicity of sodium-5-nitroguaiacolate two mechanistic profilers, DNA binding by OECD and DNA binding by OASIS, and one endpoint specific, Mutagenicity/Carcinogenicity alerts by Benigni/Bossa, were applied in the OECD QSAR Toolbox. Experimental data for genotoxicity were retrieved from two databases Carcinogenicity & Mutagenicity ISSCAN (Istituto Superiore di Sanità (ISS) a database on chemical carcinogens called ISScaN: "chemical carcinogens: structures and experimental data") and Genotoxicity OASIS.

Following the previous analysis, as would be expected, the profiling of the queried compound indicates that the genotoxicity might be due to DNA binding caused by the presence of a nitroaromatic structural feature. The OECD profiler also includes those structural alerts that might be metabolically activated to reactive electrophiles and able to interact with DNA through covalent binding. In this case, the aromatic nitro group is activated to the reactive nitrenium ion.

The next step is to define the category for the queried compound. All the steps to define the category are shown in the flow diagram in Figure 3.1.

Analogues with the *aromatic nitro compound* property were retrieved from Toolbox databases (ISSCAN and OASIS) resulting in an initial category of 791 analogues.

Sub-categorisation of the initial group of compounds using the DNA binding profilers within the OECD (Q)SAR Toolbox identified 523 analogues. This ensured that the members of the category are predicted of undergoing the same DNA binding mechanism.

Sub-categorisation with DNA binding by OASIS further decreased the number of analogues. It removed those that have also other identified DNA binding mechanisms resulting in a total of 415 analogues.

Sub-categorisation by searching for analogues with Carcinogenicity/Mutagenicity Benigni Bossa structural alerts identified 367 nitroaromatic analogues.

Empirical sub-categorisation was applied to define the structural domain. Sub-categorisation by organic functional groups (arene, ether, methyl, nitro, phenol) resulted in 42 analogues.

Further sub-categorisation, to eliminate those substances that had more than one nitro group was applied. This consisted of applying structural similarity search according to Dice similarity coefficient (atom pairs) and elimination of those compounds with similarity less than 80%. This step eliminated all those compounds that had on a benzene ring more than one nitro group. This final step resulted in 11 category members.

The flow diagram below graphically summarises the selected criteria that were in a stepwise manner applied in the QSAR OECD Toolbox for the formation of a category based on the same mechanism related to the formation of electrophilic nitrenium ion. The text in rectangles represents the criteria applied based on which specific compounds were selected. The numbers represent the retained and removed structures. The steps represented are explained in detail in the text above.

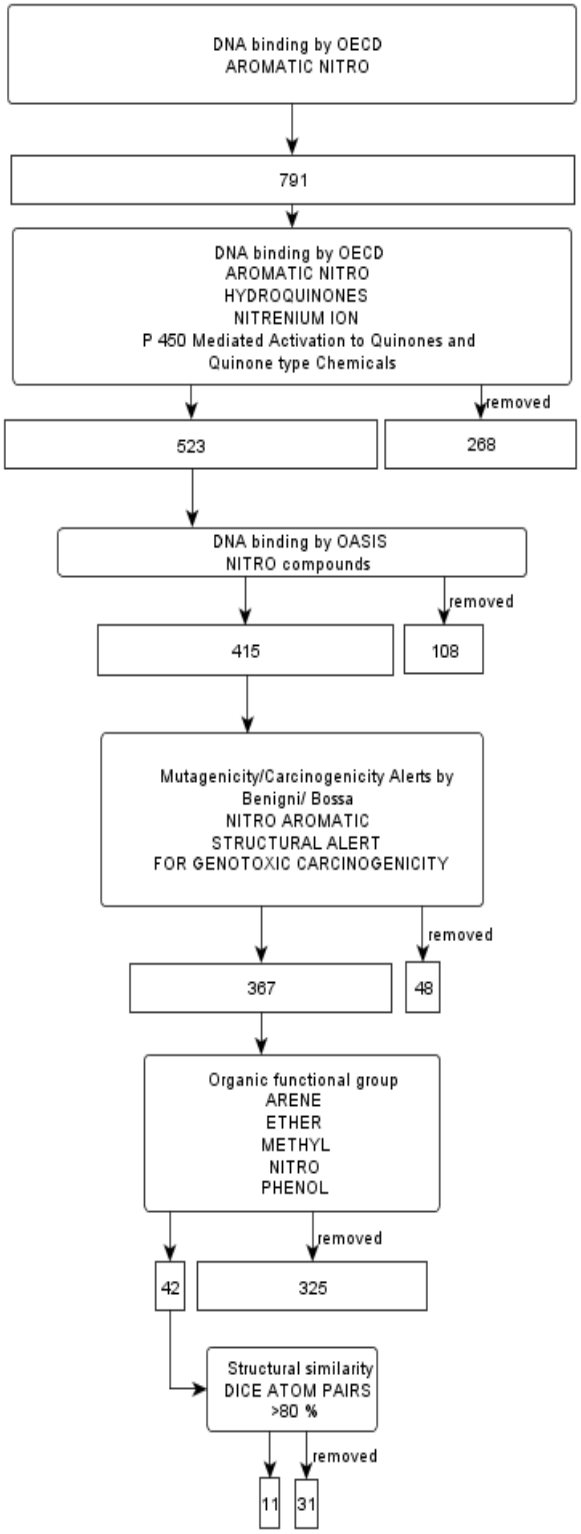


Figure 3.1. Grouping of compounds according to DNA binding via the formation of nitrenium ion as considered in the OECD QSAR Toolbox.

The final category consisted of 11 members which share the same and a single mechanistic profile. The difference between the target and the category members is that the compound of

interest has all the substituents, namely the methoxy and the hydroxy group, whereas the category members have either methoxy or hydroxy or methyl as substituents.

OECD Toolbox 2.0.27.406 (BETA); Subcategorized: Structure Threshold=50%; Dice(Atom pairs), page 1/1

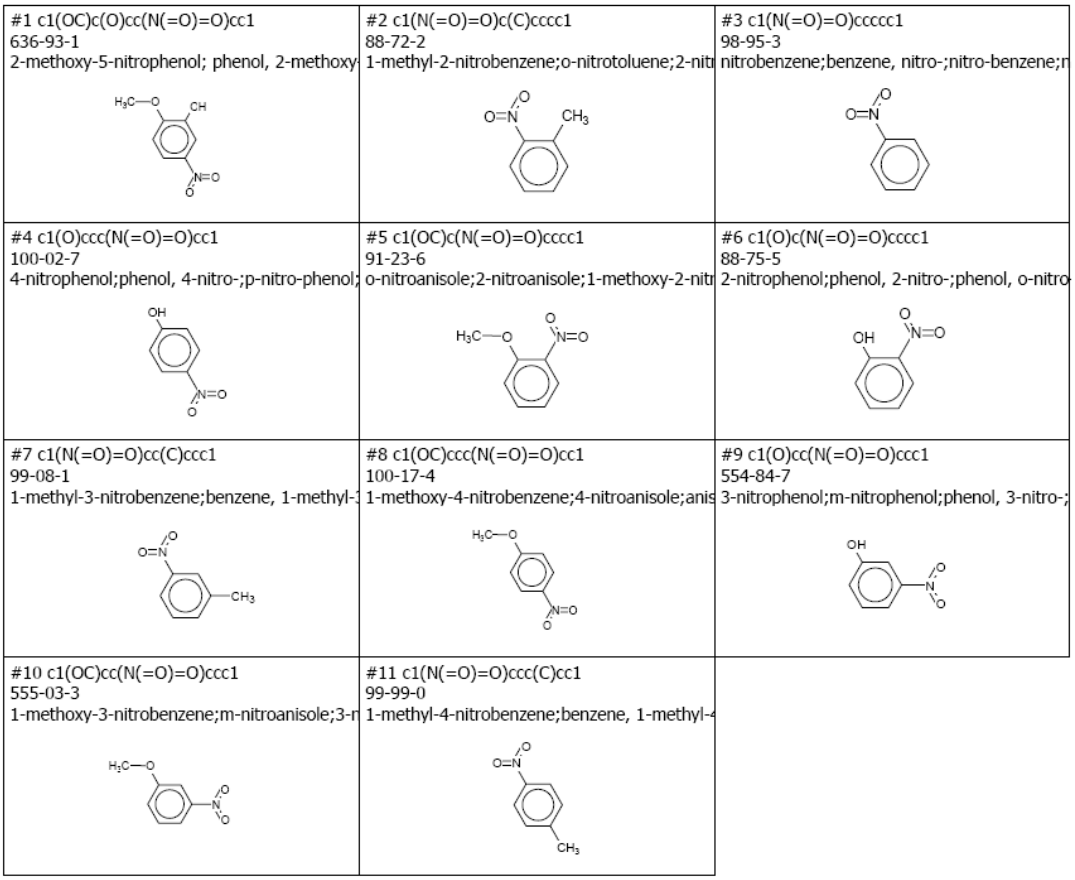


Figure 3.2. Category for the target compound considering the formation of nitrenium ion

The sub-categorisation using the mechanistic and empirical profilers resulted in a category with a well-defined mechanistic and structural domain. A group with 11 members was identified and these compounds are shown in Figure 3.2. All members of the category could potentially form a reactive nitrenium ion, which can covalently bind DNA. This category was used to fill the data gap that is present for bacterial reverse mutation (Ames test), for the query compound by read-across. The read-across prediction is based on the analysis of the distribution of the nearest five neighbours in terms of hydrophobicity (i.e. on their log P values).

Figure 3.3 shows the distribution of the category vs log P (as presented in the OECD QSAR Toolbox). It is predicted that the target compound (circled dot) would be negative, if tested for

bacterial reverse mutation without liver extract homogenate (i.e. in the absence of metabolic activation). Inspection of the five nearest chemicals shows two compounds, 4-nitroanisole and 3–nitrophenol, both have positive and negative experimental data. This could be due to inter-laboratory reproducibility of the experiment or due to different strains of *Salmonella typhimurium* used for testing.

It is also predicted that the target compound (circled dot) would be negative, if tested for bacterial reverse mutation with liver extract homogenate (Figure 3.4) (i.e. in the presence of metabolic activation). In this case, all the five nearest chemicals have negative experimental values.

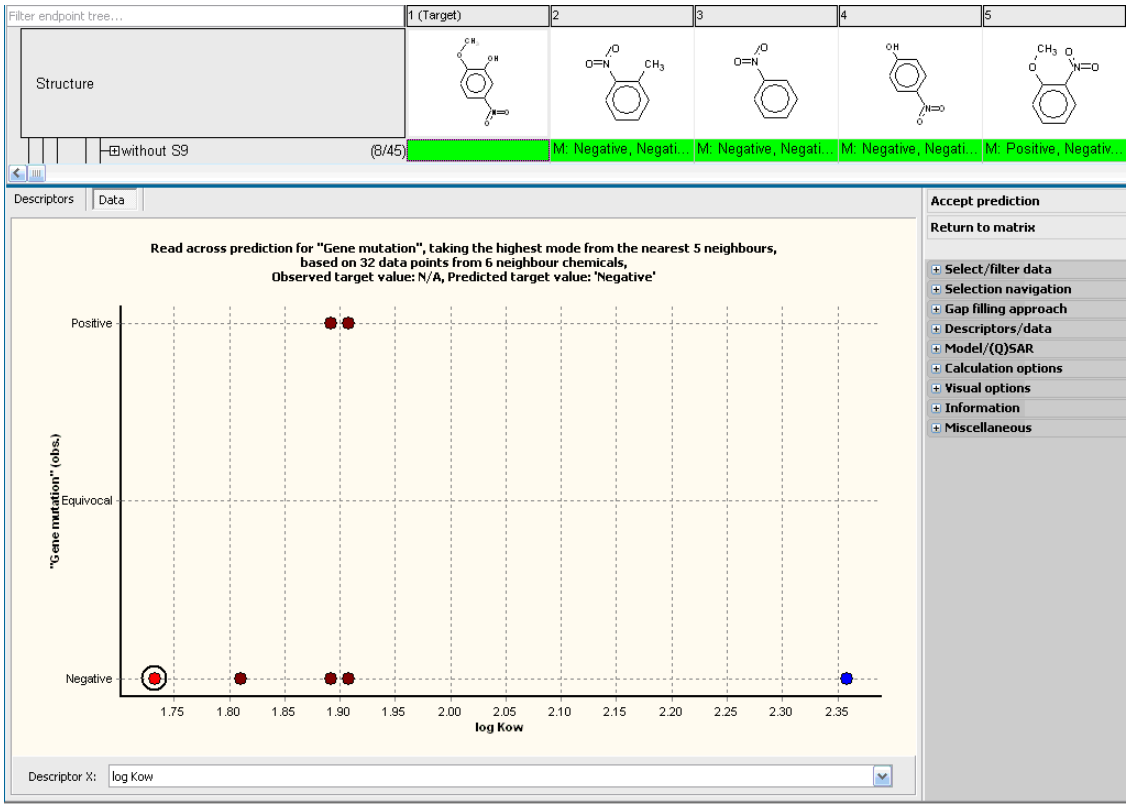


Figure 3.3. Read-across to fill the data gap in Ames genotoxicity without liver homogenate for sodium 5-nitroguaiacolate.

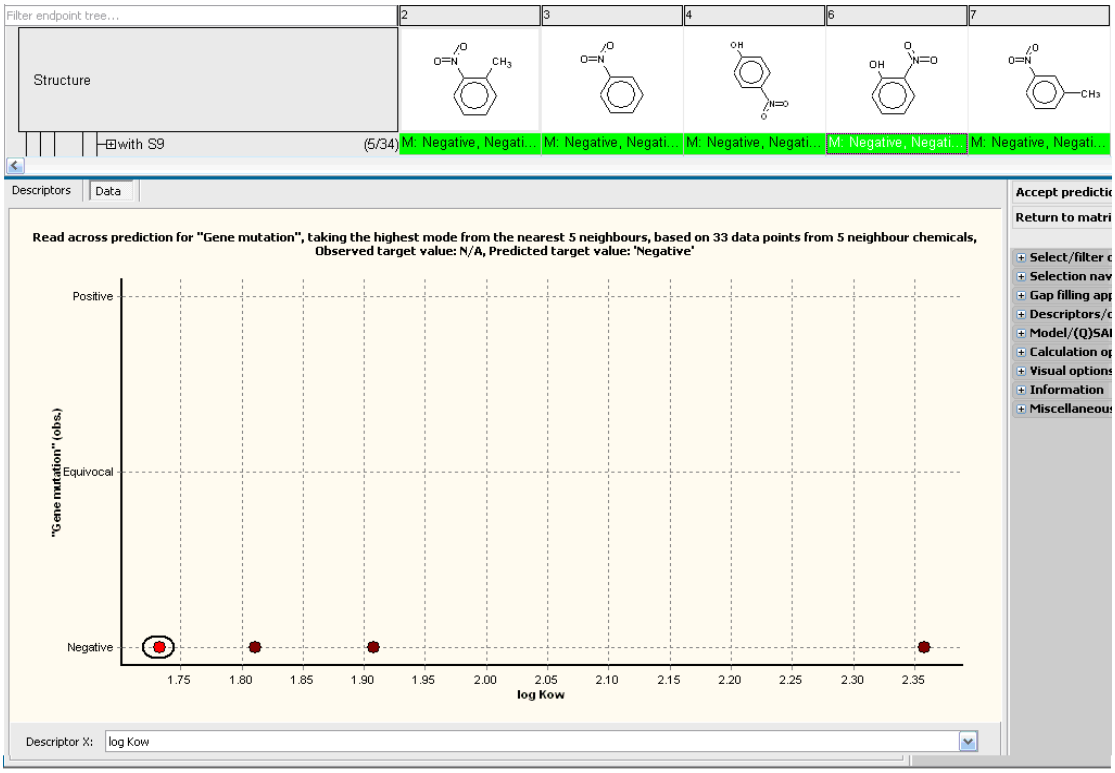


Figure 3.4. Read-across to predict Ames genotoxicity with liver homogenate for sodium 5-nitroguaiacolate

Following grouping, the target compound was predicted to be negative, for bacterial reverse mutation with, and without, the liver homogenate. This evaluation is in line with the experimental values as reported in the EFSA human health assessment for the pesticides sodium 5-nitroguaiacolate, sodium *o*-nitrophenolate and sodium *p*-nitrophenolate (EFSA, 2008). It can be concluded that although the chemical contains the structural alert for genotoxicity it is predicted to be negative for reverse bacterial mutation – Ames genotoxicity. The negative Ames results for the compound under study can be discussed in the context of possible structural features that prevent DNA reactivity, these are the mitigation factors, which include steric hindrance that blocks the attack on the electrophilic site, electronic deactivation of the electrophilic centre and include also increased detoxicifiaction (Enoch and Cronin 2010).

By looking at the members of the category based on which the Ames test results are predicted, it could be speculated that may be the hydroxyl substituent has an impact on reactivity or it might be that simple aromatic nitrobezenes are easily metabolised.

One criticism to the predictivity of Ames test it is that it is not most predictive for rodent carcinogenicity (Kirkland, Aardema, Henderson, & Müller, 2005). Prokaryotic cells lack receptors that are required by some carcinogenic chemicals for their mode of action, and those that cause large DNA deletions (would delete the target gene thus not allowing reverse mutation) would not be expected to mutate Ames bacteria. As a result of this, it could be concluded that this might not be the case for the studied compound as carcinogenicity test was negative. A way forward, to investigate the possible cause of negative Ames results could include the stability of formed metabolites as unstable metabolites could possibly explain a negative Ames.

3.4.2.1 Binding to DNA Through the Formation of Quinones

While profiling 5-nitroguaiacolate with the OECD QSAR Toolbox another mechanism involved in the formation of DNA adduct was predicted. The compound was considered to belong to the category of hydroquinones and thus the DNA binding through the formation of quinones was assumed (Figure 3.5). Quinones are Michael acceptors that alkylate cellular proteins and DNA resulting in cellular damage and DNA adduct formation. Hydroquinones have been shown to be oxidised to quinones, whereas methoxy quinones undergo demethylation to produce the corresponding hydroquinones that are then rapidly oxidised to form quinones. The reaction is a cytochrome P-450 mediated oxidative *O*-dealkylation reaction (Kalgutkar 2005, Enoch 2010). The electrophilic nature of quinones is cited as the most likely explanation for their toxicity.

Following this mechanistic consideration a category of compounds undergoing this specific mechanism was populated and then the read-across was used to fill the data gap for genotoxicity.

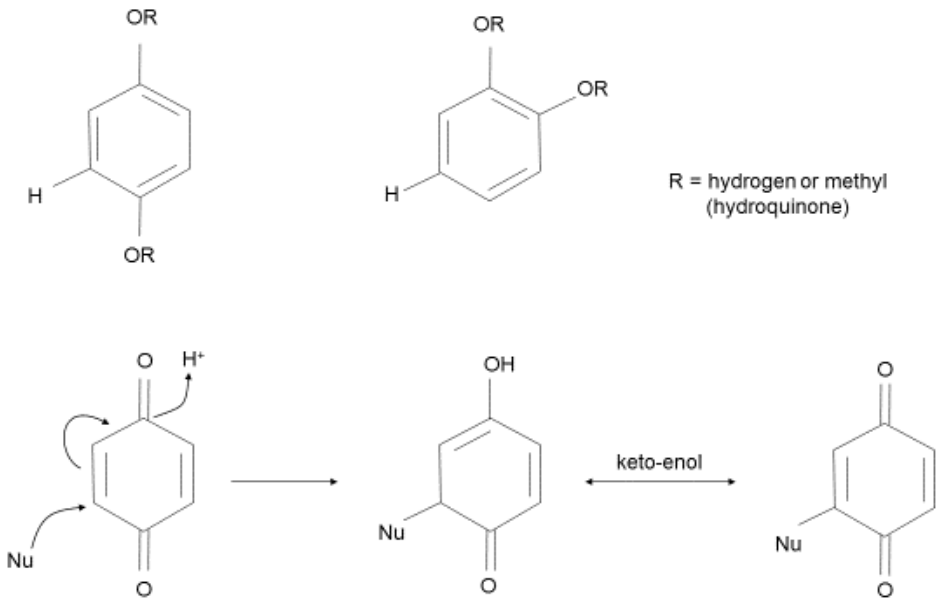


Figure 3.5. Binding to cellular nucleophiles through the formation of quinines adopted from Enoch and Cronin 2010

Grouping compounds according to DNA binding via the formation of quinones followed similar steps as applied to the grouping via nitrenium ion formation as discussed above. To summarise, the category definition followed the steps shown in Figure 3.6:

1. The information that the compound belongs to the category of hydroquinones was used to retrieve analogues from the Toolbox databases mentioned above (DNA binding by OECD) (167 analogues).

2. The members of the category should be predicted to undergo the same DNA binding mechanisms, therefore sub-categorisation of the initial group of compounds using DNA binding by OECD selected 108 analogues.

3. Following sub-categorisation with organic functional groups removed chemicals with differing functional groups to those found in the target chemical. Based on this, 11 analogues were selected with the same functional groups (arene, ether, methyl, nitro, phenol).

Again the flow diagram below graphically summarises the selected criteria that were in a stepwise manner applied in the QSAR OECD Toolbox for the formation of a category based on the same mechanism related to the formation of electrophilic quinones. The text in rectangles represents the criteria applied based on which specific compounds were selected. The numbers are representing the retained and removed structures. The steps represented are explained in detail in the text above.

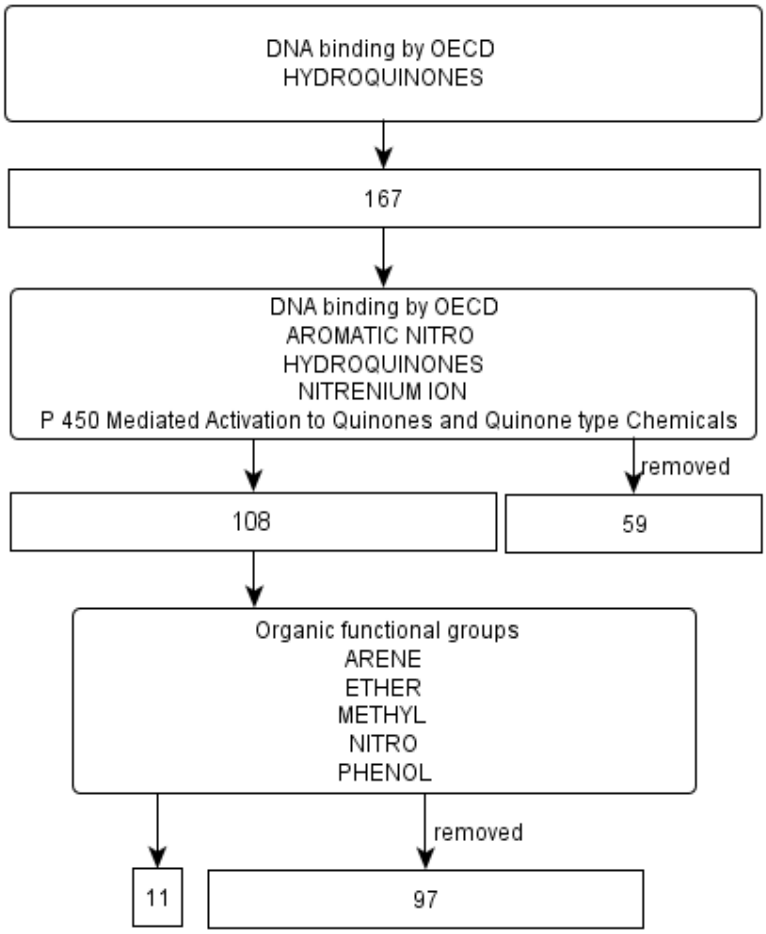


Figure 3.6. Grouping of compounds according to DNA binding via the formation of quinines for sodium 5-nitroguaiacolate.

The resulting mechanistic category, represented in Figure 3.7., included 11 category members among which a single compound was a nitro compound. This was expected as emphasis was placed on the mechanism for the formation of electrophilic quinone.

As the category formed does not represent the nitrobenzenes, there is considerable uncertainty in the application of read-across. This case study shows one of the limitations of category formation when based solely on a functional group approach to defining similarity. It provides an incentive to develop procedures of category formation where more than one mechanisms of action is predicted.

OECD Toolbox 2.0.27.406 (BETA); Subcategorized: Organic functional groups, page 1/1

#1 c1(OC)c(O)cc(N(=O)=O)cc1 636-93-1 2-methoxy-5-nitrophenol;phenol, 2-methoxy-	#2 c1(O)c(O)cccc1 120-80-9 1,2-benzenediol;catechol;pyrocatechol;pyroca	#3 C(C)(C)(C)c1c(O)ccc(OC)c1 25013-16-5 butylated hydroxyanisole;1,1-dimethylethyl-4-
#4 C(C)(C)(C)c1cc(O)c(O)cc1 98-29-3 4-tert-butylpyrocatechol;p-tert-butylcatechol;	#5 C(C)(C)(C)c1c(O)ccc(OC)c1 121-00-6 2-tert-butyl-4-methoxyphenol;phenol, 2-(1,1-	#6 c1(OC)ccc(OC)cc1 150-78-7 1,4-dimethoxybenzene;p-dimethoxybenzene;
#7 c1(O)c(O)cc(C)cc1 452-86-8 4-methyl-1,2-benzenediol;4-methylbenzene-1	#8 c1(O)c(O)c(OC)ccc1 934-00-9 3-methoxybenzene-1,2-diol;3-methoxycatech	#9 c1(O)ccc(OC)cc1 150-76-5 4-methoxyphenol;phenol, 4-methoxy-;4-meth
#10 c1(O)c(OC)cccc1 90-05-1 2-methoxyphenol;guaiacol;4-methylcatechol;	#11 c1(O)c(O)c(O)ccc1 87-66-1 1,2,3-benzenetriol;pyrogallol;1,2,3-trihydroxy	#12 C(C)(C)(C)c1c(O)c(N(=O)=O)cc(OC)c1 59282-34-7 -

Figure 3.7. Category for the target compound considering the formation of a quinine for sodium 5-nitroguaiacolate.

The read-across prediction is based both on experimental data with the liver homogenate (Figure 3.8) and without (Figure 3.9) the liver homogenate. In both cases, the predicted target (circled dot) value is negative. Inspection of the results shows that in none of the cases were any positive results observed. It can be concluded that the mechanism of DNA binding through the formation

of quinones in the case of sodium 5-nitroguaiacolate would not take place and the structural alert does not raise concern for potential genotoxicity.

It should be noted that the predictions of DNA binding are not applicable, for example, for regulatory purposes as the category based on which genotoxicity is predicted, does not include the chemicals with nitro substituent on the benzene ring.

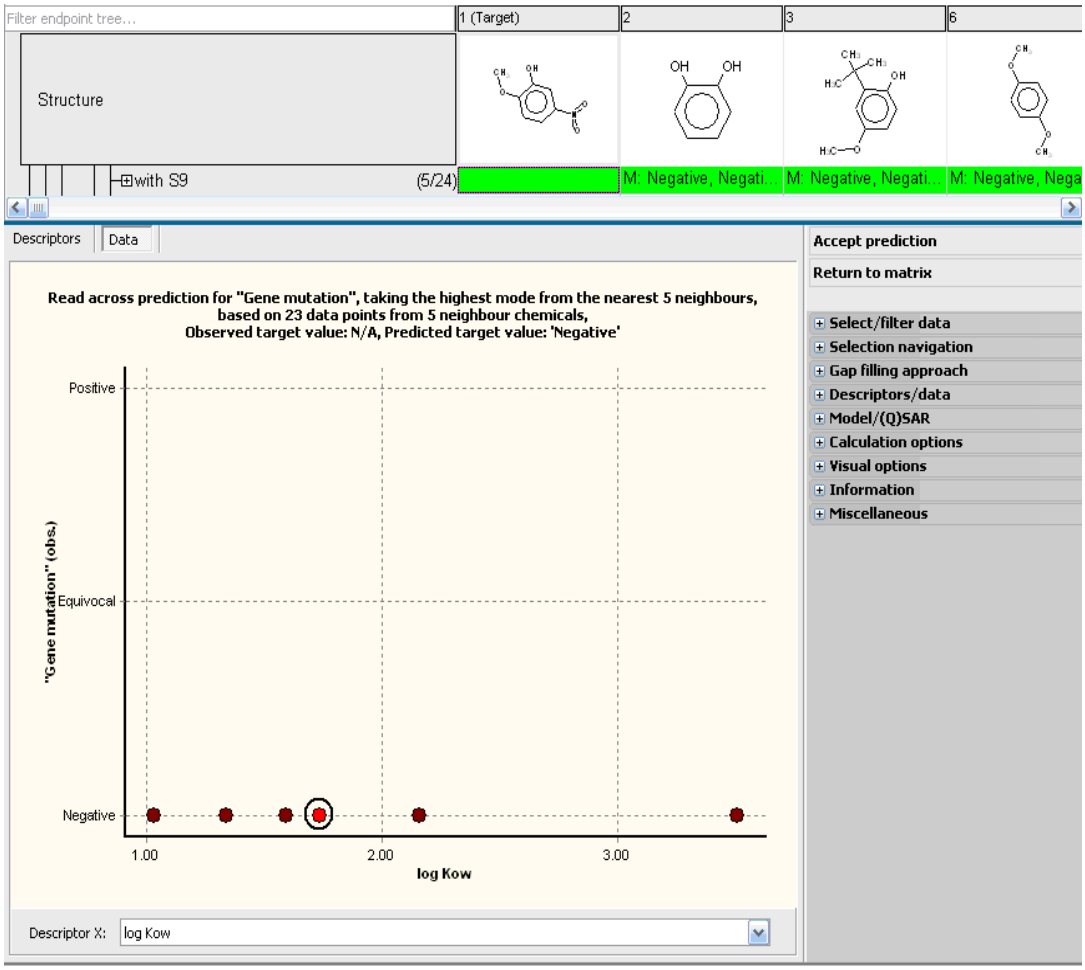


Figure 3.8. Read-across to predict Ames genotoxicity with liver homogenate for sodium 5-nitroguaiacolate.

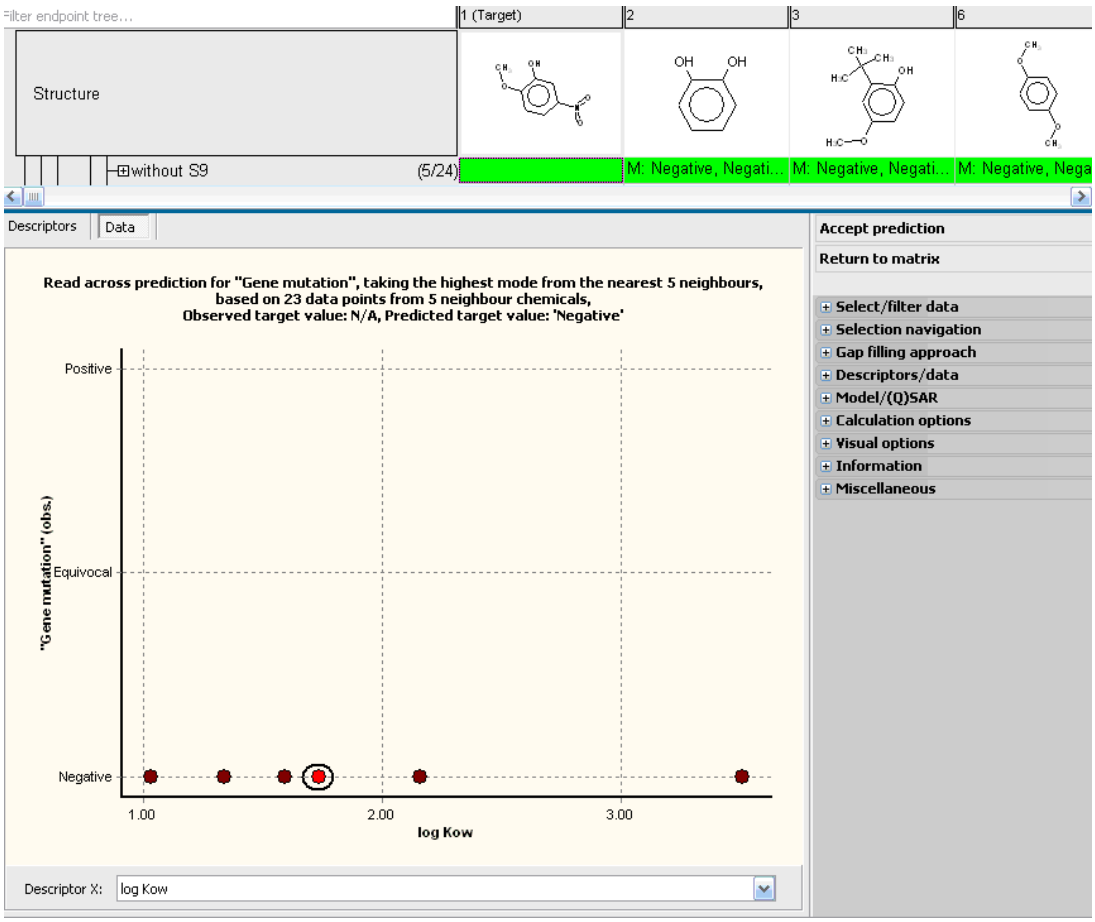


Figure 3.9. Read-across to predict Ames genotoxicity without liver homogenate for sodium 5-nitroguaiacolate.

3.5 Discussion

The aim of this chapter was to investigate specific example of how computational models could assist in the application of TTC when it comes to identifying a potential genotoxicity structural alert in a chemical. The example goes beyond identifying the structural alert as it also tries to predict genotoxicity following category formation by applying read-across. The computational toxicology tools applied were able to identify a structural alert for genotoxicity and predicted sodium 5-nitroguaiacolate as genotoxic.

The extended decision tree for the application of TTC (Felter et al., 2009), which considers the availability of Ames data, was also investigated. When experimental data are lacking these can be evaluated by chemical grouping and read-across, which is recognised as an important

approach for filling data gaps (Ball et al., 2016). Populating a category with structures that have a common mechanism of action and share the same structural features allowed prediction of Ames genotoxicity for the compound under investigation. It was predicted to be negative. Application of this selection of tools, combination of QSAR models and category approach and read-across in the case of sodium 5-nitroguaiacolate proved efficient in identifying a structural alert and predicting negative Ames results.

For the application of the TTC decision tree, according to the recent review of the TTC by EFSA and WHO (EFSA & WHO, 2016b), if non-genotoxicity is demonstrated based on an WoE approach, a substance would not to be expected to be of safety concern. In the review, it was also concluded, that non-DNA reactive chemicals do have a threshold mode of action and higher TTC values are applicable. Therefore the assessment can proceed down the decision tree.

As (Q)SAR methods are intended to be part of a Weight of Evidence approach, long-term studies also were considered, which could confirm or reject the occurrence of cancer. In the case of sodium 5-nitroguaiacolate the results from long term carcinogenicity testing were negative (EFSA 2008) therefore supporting the negative results from genotoxicity predictions.

Chemicals are being developed for use in food, pharmaceuticals, plant protection products, biocides etc. Where exposure is expected to be negligible such as for additives, impurities and/or metabolites they can be assessed by the application of TTC. TTC relies on the identification of a SA which, as demonstrated, can be identified by the application of *in silico* tools. An identified SA for genotoxicity might not be sufficient to raise concern, therefore the category approach and read-across provides viable option(s) to fill a data gap for this toxicological endpoint.

4 Application of TTC to Cosmetic Ingredients

4.1 Introduction

A wide range of cosmetic products is used daily by consumers from all age groups. Cosmetic products are used for daily personal hygiene such as toothpaste, soap and shampoo and for beauty purposes, such as products including hair dyes, makeup and perfumes for example. Due to a variety of products, diversity in exposure scenarios is expected, such as exposure from cosmetic products applied to the skin to be immediately washed off, exposure from products that undergo a chemical reaction on hair, exposure from leave-on products or exposure from products meant only for children or pregnant women.

Cosmetic products are made of ingredients among which some are recognised to pose health risks, therefore some ingredients are prohibited for use in cosmetics and the use of some ingredients is restricted by setting maximum concentration levels in ready-to-use products. For example, some phthalates and parabens are prohibited and the use of boric acid and ammonia is restricted. Among the hazard effects associated with cosmetic ingredients are local toxicological effects such as skin, eye and respiratory irritation, sensitisation and systemic effects such as carcinogenicity, mutagenicity and reproductive (CMR substances) toxicity. In the EU, CMR substances have been prohibited as cosmetic ingredients from December 1st, 2010. Among prohibited substances are, for example, dibutyltin dichloride, maneb and mancozeb. Further, in the EU, the *Regulation (EC) No 1223/2009 on Cosmetic Products* which regulates cosmetic products placed on the market, and is discussed in the introductory chapter, provides incentives to the research presented in this chapter.

4.1.1 The COSMOS Project and the Threshold of Toxicological Concern

Under the Integrated *In Silico* Models for the Prediction of Human Repeated Dose Toxicity of COSMetics to Optimise Safety (COSMOS) Project, tools and workflows were developed to enable reliable estimate of repeat dose toxicity to humans for cosmetic ingredients and their final products. One key aspect of COSMOS activities included the concept of Threshold of

Toxicological Concern (TTC), which also represents an alternative approach to animal testing (Yang et al., 2014).

The TTC concept applies thresholds for three structural classes of chemicals based on the Cramer classification (Cramer & Ford, 1978). This was subsequently used by Munro and co-workers with the purpose of deriving human exposure levels (TTC values) based on structure and oral systemic toxicity data for toxicity endpoints other than carcinogenicity (Munro et al., 1996). A detailed description of the Cramer classification scheme and the derivation of exposure threshold values are given in the introductory chapter.

4.1.1.1 The concept of Threshold of Toxicological Concern in the Context of Cosmetic Products

TTC is a risk assessment tool where exposure level for a chemical is established below which there is negligible risk to human health. According to the TTC concept, identification of a safe level of exposure to a chemical with unknown toxicological profile, based on its structure and toxicity data of similar compounds, is possible. Therefore, in order to estimate the toxicological potency of chemicals with unknown toxicological profile, TTC relies heavily on chemical structure, existing toxicity data and exposure data.

TTC was originally developed to assess food additives and food flavourings, meaning exposure through food intake. However, it has also been evaluated for implementation in other regulatory areas such as personal care products. Blackburn et al. (Blackburn et al., 2005) concluded that the NOEL values for ingredient chemicals were in the same range of the original Munro's dataset, thus supporting the currently available TTC for use in personal care products.

One important aspect of the original TTC concept is that it is based on oral exposure data, whereas cosmetic ingredients application is mostly topical. A study by Kroes et al. (Kroes et al., 2007) regarding the application of TTC to cosmetics concluded that the application of the TTC approach to cosmetic ingredients is valid and that TTC values are appropriate for the safety evaluation of systemic exposures resulting from topical applications of cosmetic ingredients.

TTC thresholds are based on oral exposure, however cosmetic ingredients are applied topically. Therefore, the application of conservative default uncertainty factors should be considered in order to compensate for route dependent differences.

An issue arising from the specific use of cosmetic products and the application of TTC is that current threshold values may be protective for systemic toxicity through oral exposure but may not be protective for local effects such as skin sensitisation. Safford et al. addressed this issue in 2008 (Safford, 2008) and later in 2015 (Safford et al., 2015) where the possibility of establishing a dermal sensitisation threshold (DST) similar to TTC is supported, including for protein reactive chemicals.

Based on the preliminary analysis report, produced by the JRC (Worth, Cronin, Enoch, & Fioravanzo, 2012) in the context of application of TTC to cosmetics, the three scientific committees (SCCS/SCHER/SCENIHR) of the European Commission issued an opinion on the use of TTC for human safety assessment of chemical substances with a focus on cosmetics and consumer products. They considered that the TTC is scientifically acceptable; nevertheless, this is a probability-based screening tool and may have additional uncertainties, because the derivation is based on frequency distribution and the TTC value selected is not the lowest value but is a value close to the lowest point. Following the probability of a cancer risk higher than 10^{-6} above background levels is estimated to lay between 0 and 5%. The Opinion also identifies that for some chemical structures the underlying database is not adequately supportive which is also true for some toxicological endpoints, such as local effects. Due the vast application of cosmetic products, appropriate exposure assessment, to cover all scenarios, is a prerequisite in the application of the TTC approach for cosmetic ingredients.

4.2 Aim of the Study

The aim of this study was to assess the representativeness of the underlying Munro non-cancer TTC for the chemical space of cosmetic ingredients and further to evaluate the degree of protectiveness provided by Cramer class-related Munro threshold values for cosmetic ingredients.

4.3 Datasets and Methods

4.3.1 The COSMOS TTC Dataset

Through the COSMOS Project a non-cancer TTC COSMOS dataset was developed by matching cosmetics ingredients in the Cosmetic Inventory with oral repeat dose toxicity data from five toxicity data sources: The Munro dataset, RepDose, ToxRefDB, FDA PAFA and ILSI DevTox (Worth et al., 2012). The present author was not involved in extraction of NOEL values for cosmetic ingredients as these were mainly harvested by a COSMOS Project partner Altamira LLC. The summary of criteria for the selection of NOELs from various data sources was as follows (Worth et al., 2012):

- Oral repeated dose toxicity studies included sub-chronic, chronic, reproductive, developmental, multigenerational reproductive-developmental, immunology, and neurotoxicity. For target organ studies, rat, mouse, dog and monkey studies were used. For reproductive and developmental toxicity, rat, mouse and rabbit studies were used.

- In general, minimum NOEL values (such as ToxRefDB or FDA PAFA) were selected. However, NOAEL values from regulatory sources were used whenever available.

Version 1.0 of the COSMOS TTC dataset consisted of 660 substances. After removal of chemicals with undefined structures, the dataset consisted of 558 substances with well-defined structures. For the purposes of the Cramer classification and NOEL distribution analyses in this study, the dataset was further reduced to 385 substances with well-defined structures (excluding inorganics, organometalics, polymers and substances for which the tested form was unknown) and with the NOEL values available for developmental and reproductive toxicity and repeat dose toxicity studies including and above 90 day duration (subchronic study).

The COSMOS Project was finalised in December 2015 and within the context of the project, the dataset has been updated in terms of its chemical coverage (adding chemicals from additional toxicity data sources) as well as the quality control of the chemical structures and toxicity data.

The data were extracted from the datasets described below as they were available in 2011. An assessment of the overlap in toxicological data from different datasets was assessed by other partners in the COSMOS Project (Hollnagel et al., 2016). It should be stressed that the author of the thesis was not involved in the extraction of toxicological data nor in the evaluation of the overlap of the datasets.

The Munro dataset is currently the *de facto* TTC database for non-cancer endpoints. The dataset, published by Munro et al (Munro et al., 1996), contains data for 613 tested chemicals. The structure data file and summary data tables are downloadable from EFSA (http://www.efsa.europa.eu/de/supporting/pub/159e.htm). The summarised data include study design, NOEL values and the associated critical effects. These files were compiled for the purposes of a chemoinformatics investigation carried out by Soluzioni Informatiche srl (Bassan, Fioravanzo, Pavan, & Stocchero, 2011) under the terms of an EFSA contract.

RepDose DB (repeated dose toxicity) is a relational database on subacute to chronic toxicity studies and at the moment of the analysis contained 655 chemicals. The toxicity of these chemicals is documented in about 2280 studies, carried out in rats, mice or dogs with oral or inhalation exposure. The database has been developed by Fraunhofer Institute for Toxicology and Experimental Medicine (ITEM) (Bitsch et al., 2006) with funding from Cefic LRI and grows continuously. It is accessible through the following web link http://www.fraunhofer-repdose.de/ upon registration.

ToxRefDB (Toxicity Reference Database) is a database containing the results of animal toxicity studies. The database is a joint effort between the United States National Centre for Computational Toxicology (NCCT) and *Environmental Protection Agency's (US EPA)* Office

of Pesticide Programs (OPP). At this time, ToxRefDB includes rat chronic and cancer assays, mouse cancer assays, rat 90-day toxicity studies, rat multigenerational reproduction studies, and rodent and rabbit prenatal developmental studies.

FDA PAFA, the Priority-based Assessment of Food Additives (PAFA) database. PAFA contains administrative, chemical and toxicological information on over 2000 substances directly added to food, including substances regulated by the U.S. Food and Drug Administration (FDA) and Generally Recognized As Safe (GRAS) and prior-sanctioned substances. In addition, the database contains only administrative and chemical information on less than 1000 such substances. The more than 3000 total substances together comprise an inventory often referred to as *"Everything" Added to Food in the United States* (EAFUS).

ILSI DevTox, International Life Sciences Institute Developmental Toxicity Database[2]. The data contained in the database have been compiled for a range of laboratory animals (rat, mouse, rabbit, and hamster) and routes of application (oral gavage, feed or water, inhalation, dermal, injection). The first version of the database contains data from 315 individual experiments, covering 166 publications and 195 different substances. The database was developed with the purpose to increase the availability of high quality toxicological data to support the development of statistically based Structure-Activity Relationship (SAR) models as tools for prioritising chemical substances for further study.

[2] http://www.ilsi.org/ResearchFoundation/RSIA/Pages/DevelopmentalToxicityDatabase.aspx

4.3.2 Methods

4.3.2.1 Cosmetics Ingredients Inventory, COSMOS TTC and Munro Structure Files

A selection of computational methods was applied in order to identify substructures, calculate physico-chemical descriptors and apply the Cramer classification. In order to run these tools, molecular structural files were generated.

The .sdf files for Cosmetics Ingredients Inventory, COSMOS TTC and MUNRO were developed within the COSMOS project. The Cosmetics Ingredients Inventory file contained 4458 chemicals, that were extracted from the EU and the US cosmetics lists, the COSMOS TTC file contained 558 well-defined chemical structures and finally, a file with 596 well defined, unique, chemical structures representing the Munro dataset.

For the Cramer classification and NOEL distribution analysis the COSMOS TTC dataset was reduced and contained 385 compounds identified with the relevant NOEL values. For the purpose of the analysis a .smi file was generated by copying the elements of the dataset into Notepad and saving the file as .smi.

4.3.2.2 Characterisation of the Chemical Space

In the first step, chemical space defined by structural features representing the datasets was characterised by evaluating the presence of Leadscope-defined chemical substructures. Molecular structural files, the .sdf files for the three datasets, namely the Cosmetics Ingredients Inventory, COSMOS TTC and Munro, were imported into the Leadscope tool v2.4 in order to identify chemical substructures. Results were stored as spreadsheet files and further evaluated and visualised with Microsoft Excel.

Another way to examine the chemical space is to evaluate the overlap in physio-chemical properties describing each set, namely the Munro and COSMOS TTC. For that purpose, key physicochemical descriptors were selected representing molecular size (weight), partitioning behaviour (log P), solubility (log S) and reactivity (polarisability). Physico-chemical descriptors

were calculated with *CORINA Symphony* (previous *Adriana.Code*) and further visualised in 3D graphs using *STATISTICA 10* software tool to enable the assessment of the overlap. With Microsoft Excel, the range of values was evaluated and averages were calculated.

The results presented are a preliminary analysis of the chemical space and it is also acknowledged that other analyses, in terms of structural features representations and physico-chemical parameters, have been performed by other partners in the COSMOS project (Worth et al., 2012). Bassan et al. (Bassan et al., 2011) carried out an extensive evaluation of the applicability of the physico-chemical data in TTC .

4.3.2.3 Calculated Molecular Properties

The physico-chemical descriptors for the present analysis were selected according the Lipinski rules.

Molecular Weight is related to the size of the molecule. As molecular size increases, solubility decreases, it also impedes passive diffusion through the bilayer membrane.

Solubility. The calculated solubility, given in mol / L [log units], represents the maximum dissolved concentration in water. It represents one of the most important molecular properties in the discovery of new active substances as low solubility limits absorption and causes low oral bioavailability.

Partitioning coefficient the partition coefficient (P) is the ratio of concentrations of a compound in the two phases of a mixture of two immiscible solvents at equilibrium. Normally one of the solvents is water while the second is hydrophobic such as octanol. Hence the partition coefficient measures both how hydrophilic ("water-loving") or hydrophobic ("water-fearing") a chemical substance is. Hydrophobic chemicals with high partition coefficients are preferentially distributed to hydrophobic compartments such as lipid bilayers of cells while hydrophilic chemicals (low partition coefficients) preferentially are found in hydrophilic compartments such as blood and the cytoplasm.

Polarisability defines how easy an electron cloud in a non-polar molecule can be distorted in the vicinity of electrically charged species, such as ions or polar molecules with dipole moments.

4.3.2.4 Cramer Classification and Derivation of The 5th Percentile NOEL

The chemicals from the TTC COSMOS dataset were classified according to Cramer classification by using the non-extended version of the Cramer decision tree implemented in the *Toxtree tool (v2.5.0)*. The NOEL distribution was further analysed and visualised with *Microsoft Excel*.

In addition, the cumulative distribution analysis for each Cramer class and further calculation of the 5th percentile NOEL to derive Cramer class-related thresholds for cosmetic ingredients was performed using *Microsoft Excel*.

4.4 Results

4.4.1 Comparison of the Dataset in Terms of Structural Features

The aim of this exercise was to identify possible differences in the substructures representing the three datasets. Structural features analysis represented in Figure 4.1 showed that Munro dataset is lacking in organometallics, silicon-containing compounds and surfactant classes. The Munro dataset is also missing acid halides, allenes, boron-containing compounds and thiocarboxylates.

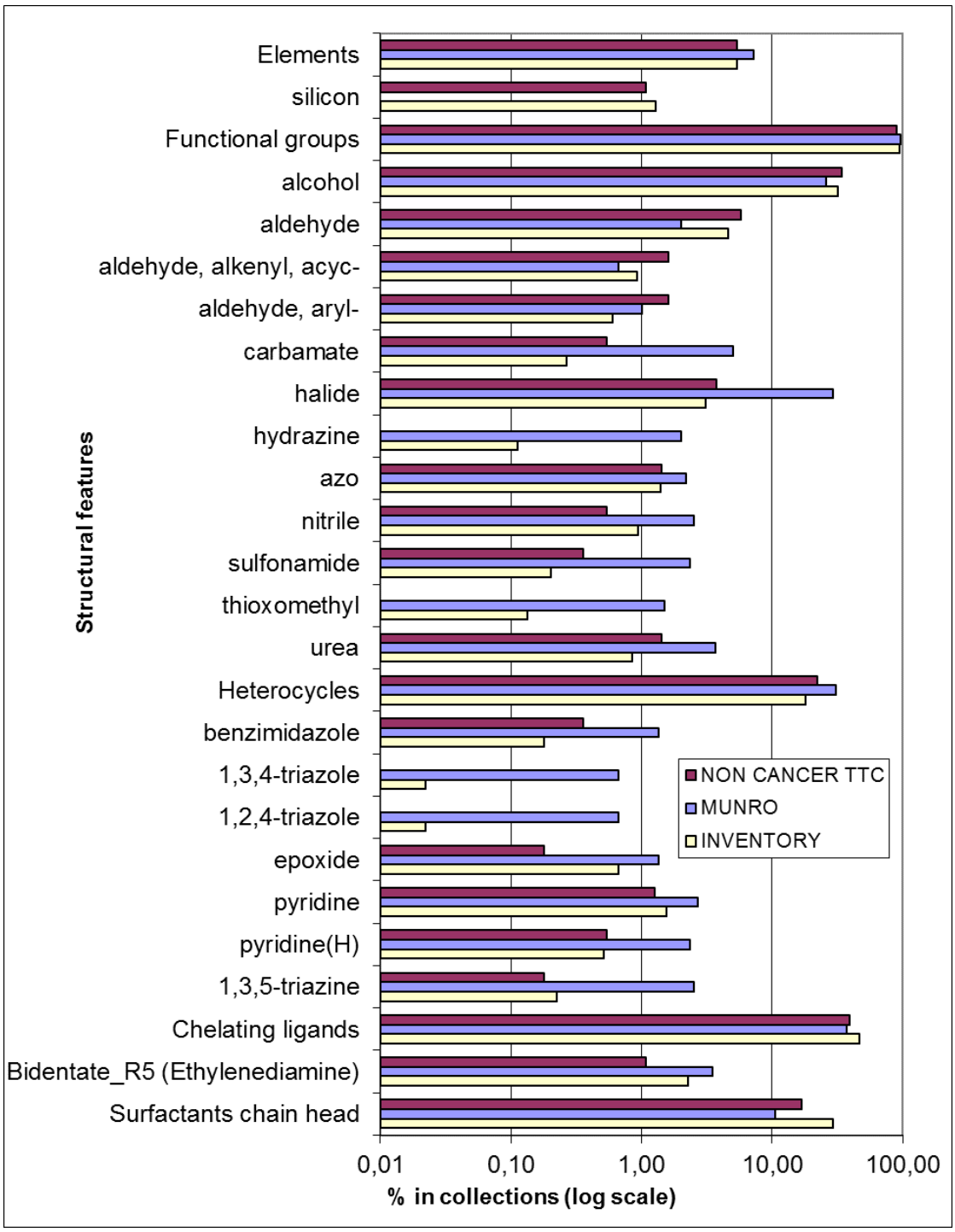


Figure 4.1. Comparison of structural features across datasets

4.4.2 Comparison of the Dataset in Terms of Physico-chemical Properties

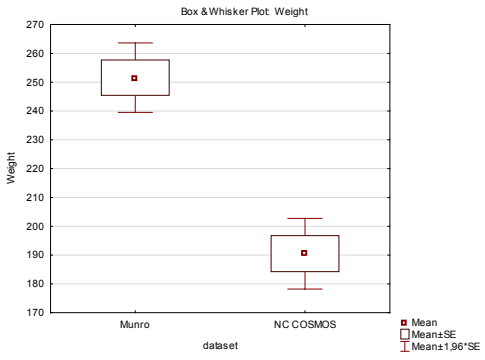
Another way to assess the overlap of the Munro and COSMOS TTC is to plot the distribution of chemical space described by calculated physio-chemical properties across three dimensions. The visualisation of distribution in Figure 4.3 and Figure 4.4 confirmed that chemical spaces

broadly overlap across the analysed descriptors, namely log S, log P, molecular weight and polarizability (Table 4.1).

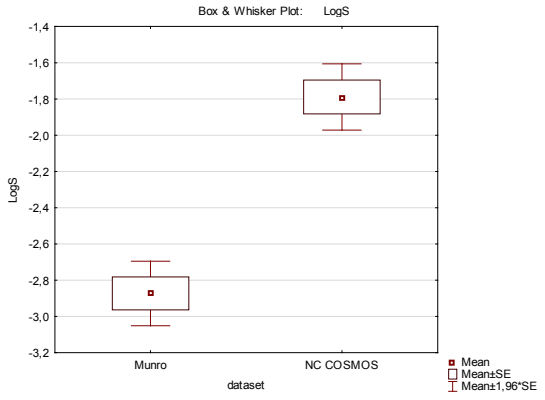
Table 4.1. Minimum, maximum and average values for calculated physicochemical properties

	TTC	MUNRO	TTC	MUNRO	TTC	MUNRO	TTC	MUNRO
	Weight		LogS		XlogP		Polariz	
Max	1701.20	1134.98	4.13	4.13	14.31	13.29	150.66	93.95
Min	17.01	30.03	-12.91	-11.52	-13.33	-13.33	1.02	2.70
Avg.	190.51	251.57	-1.79	-2.87	1.49	2.24	19.90	25.15
St. dev.	147.41	150.17	2.20	2.22	2.63	2.39	14.80	14.75

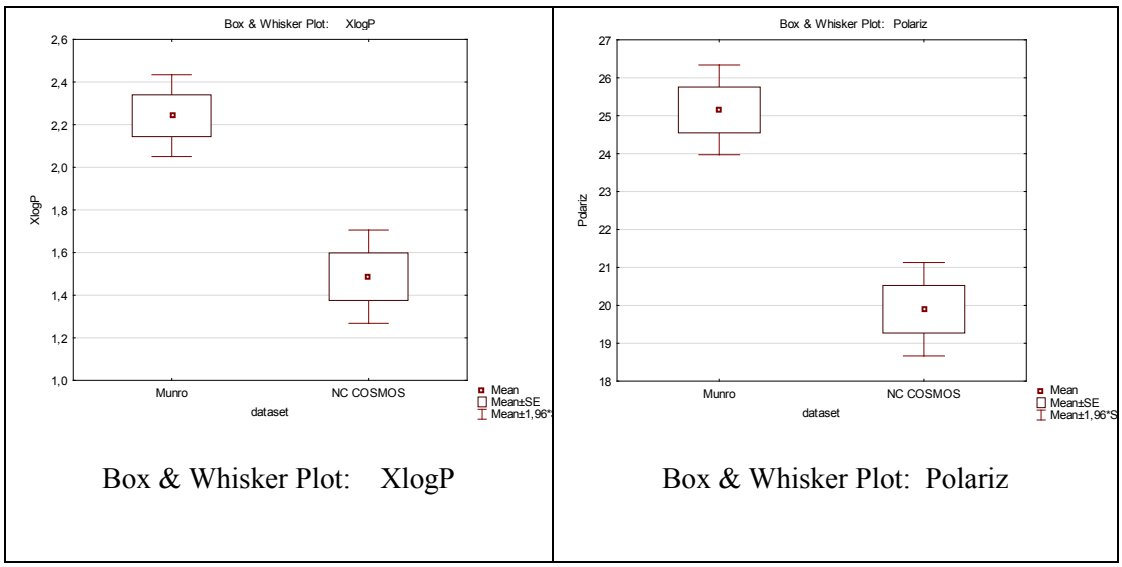
Further analysis of the values in the terms of the t-test analysis for each physico-chemical property, revealed that the COSMOS TTC dataset contains smaller structures (lower weight) than the Munro dataset and that the COSMOS TTC has higher prevalence of hydrophilic chemicals. The presented plots in Figure 4.2 show the significant differences in means of calculated properties.



Box & Whisker Plot: Weight



Box & Whisker Plot: LogS



Box & Whisker Plot: XlogP

Box & Whisker Plot: Polariz

Figure 4.2. Box & Whisker Plot for calculated physico-chemical parameters

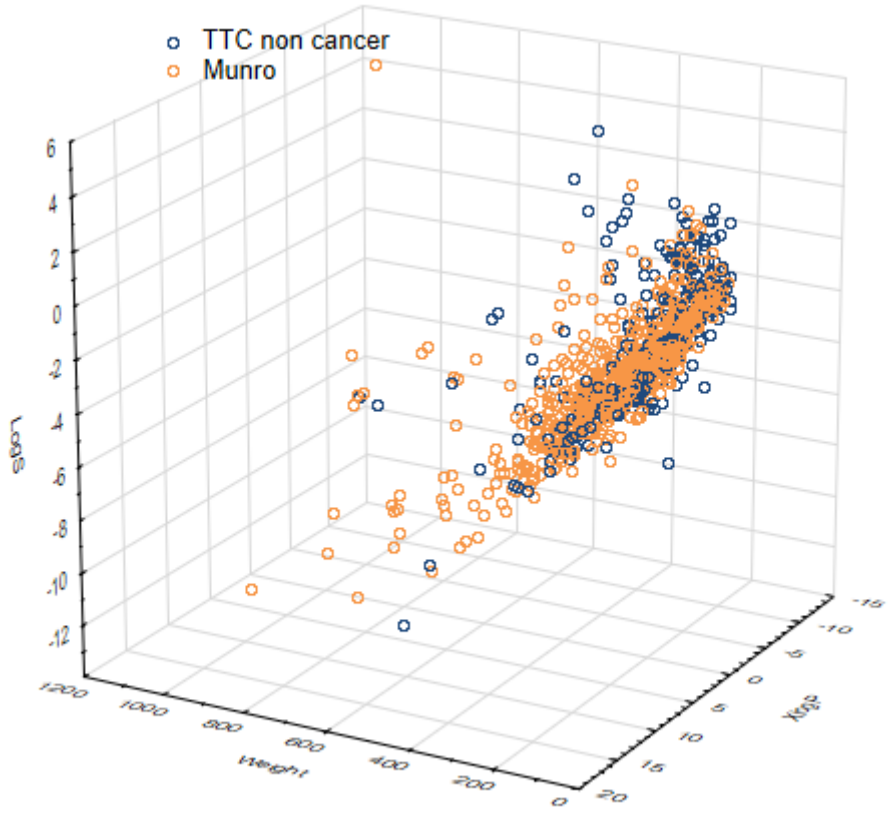


Figure 4.3. Graphical representation of the distribution of partitioning, solubility and molecular weight for the TTC non-cancer and Munro dataset.

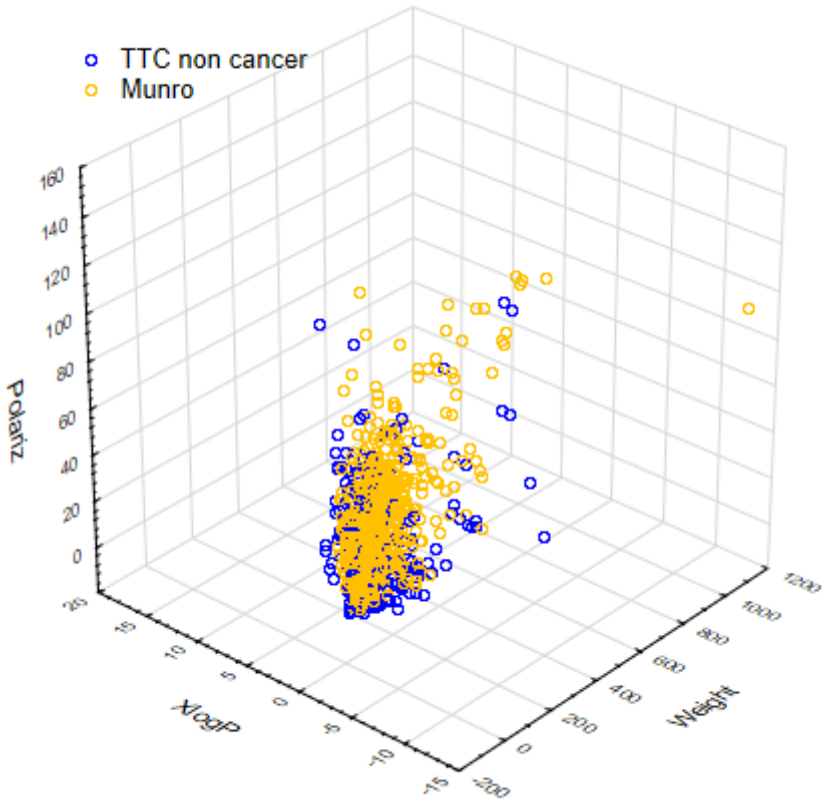


Figure 4.4. Graphical representation the distribution (weight, partitioning, polarizability) for the TTC non-cancer and Munro dataset

4.4.3 Cramer Analysis

The chemicals in the TTC COSMOS dataset were categorised according to Cramer classification by using the non-extended version of the Cramer tree (v 2.5.0). The results (Table 4.2) for the TTC COSMOS and Munro dataset are illustrated in Figure 4.5 and Figure 4.6. In the Munro dataset, most of the chemicals (75%) are in Cramer class III with a much lower proportion (21%) being found in Cramer class I. Whereas in the Cosmetics inventory and the TTC COSMOS dataset there is a fairly even balance between Class III and Class I chemicals. The percentage of chemicals classified as Cramer class II is less than 10%, which is typical of many datasets. It has been observed several times that the underlying database for Class II compounds is not well represented, making the derivation of the TTC for this potency class questionable. Therefore, Class II substances were excluded from further analysis.

Table 4.2. Cramer classification of the Cosmetics Inventory, Munro and COSMOS TTC dataset

Cramer Class	Munro	COSMOS TTC	Cosmetics Inventory
Class I	123 (21%)	201 (52%)	1977 (45%)
Class II	24 (4%)	34 (9%)	327 (7%)
Class III	449 (75%)	150 (39%)	2154 (48%)
TOTAL	596	385	4458

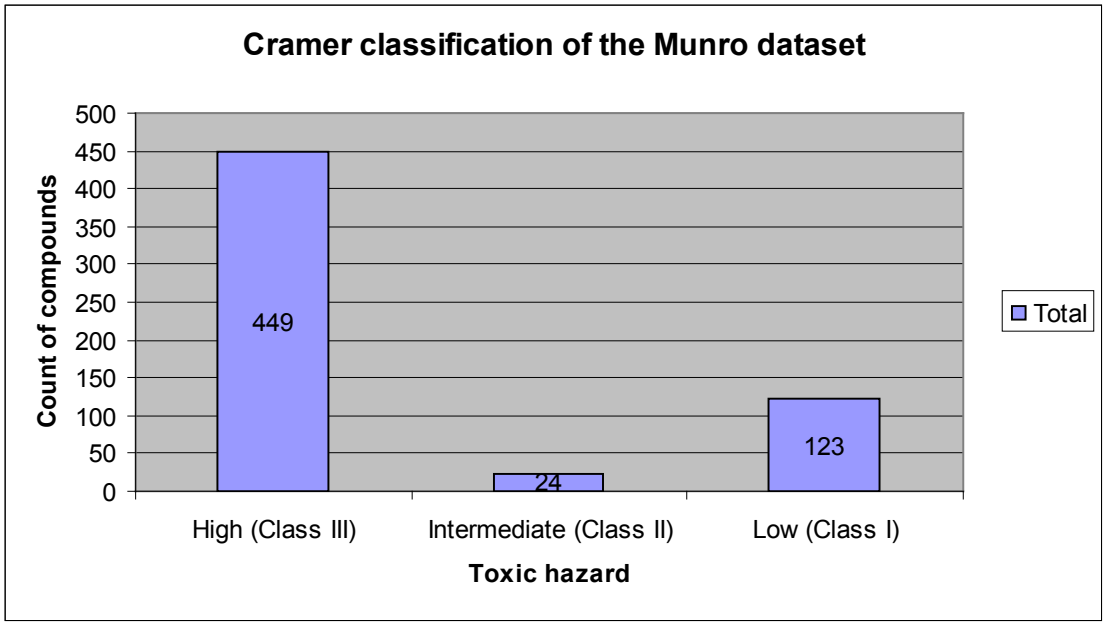


Figure 4.5. Cramer classification for the non-cancer Munro dataset

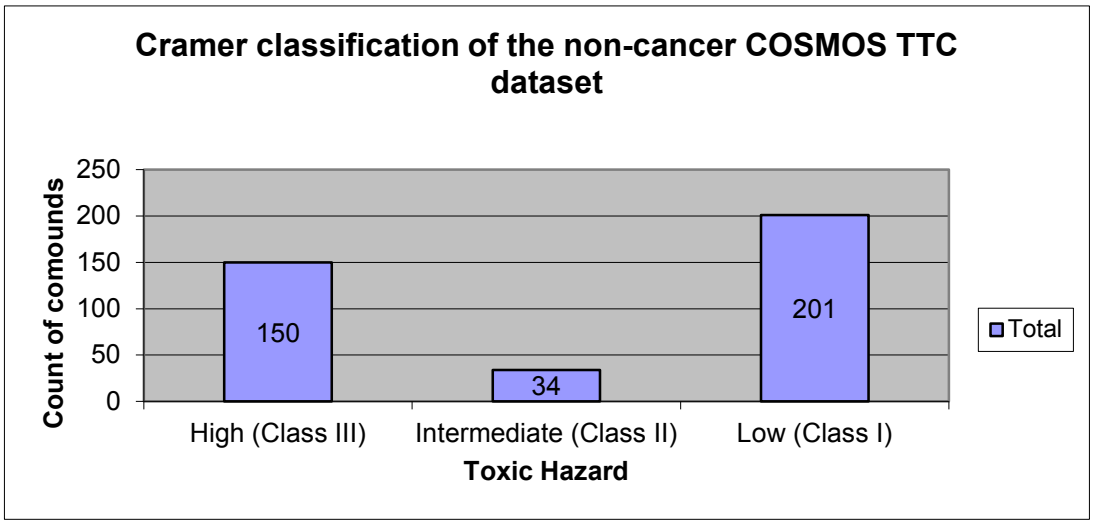


Figure 4.6. Cramer classification for the non-cancer COSMOS TTC dataset

4.4.4 Analysis of NOEL Distributions in the COSMOS TTC Datasets

To assess the degree of protectiveness provided by the Cramer-related (Munro) threshold values for cosmetic ingredients, the Munro threshold values were compared with the corresponding thresholds derived from a cumulative distribution analysis of NOEL values in the COSMOS TTC dataset. As mentioned above, this dataset is derived from multiple data sources (Munro, PAFA, ToxRefDB and RepDose) and has been subject to ongoing extension and revision within COSMOS.

The distribution analysis was applied to the lowest NOEL value for each substance in the dataset, which may be lower than the NOAEL, i.e. the lowest NOEL for a toxicologically relevant effect. Indeed, the presence of freestanding NOELs may result in an over-conservative estimate of the 5th percentiles. This is because, the free standing NOEL values are derived from studies that did not show any effects at the highest dose tested, hence these values may be very conservative and may fall far below the threshold region of the dose–response curve (Blackburn et al 2005).

Distribution analysis was performed for NOEL values of 385 structurally well-defined substances excluding inorganics, organometallics, polymers, and substances for which the tested form was unknown. This analysis included developmental and reproductive toxicity studies, but excluded all repeat dose studies with an exposure duration less than a sub-chronic study (typically 90 days). The NOEL values from sub-chronic studies were divided by a factor of 3 (Munro adjustment factor for sub-chronic to chronic conversion). The 5th percentile NOEL for the substances in each Cramer class was calculated from a theoretical log-normal cumulative distribution. The three cumulative distribution curves (Figure 4.7) clearly fit the log-normal distribution, therefore the calculations of the 5th percentile for each Cramer class was calculated according the log-normal distribution.

The 5th percentiles for the substances in each Cramer class are summarised in Table 4.3. It can be seen that in the case of Cramer Class I, the 5th percentile derived from the COSMOS dataset

(1362 µg/kg/day) is lower than the corresponding Munro value (3000 µg/kg/day)[3] by a factor of about 2.

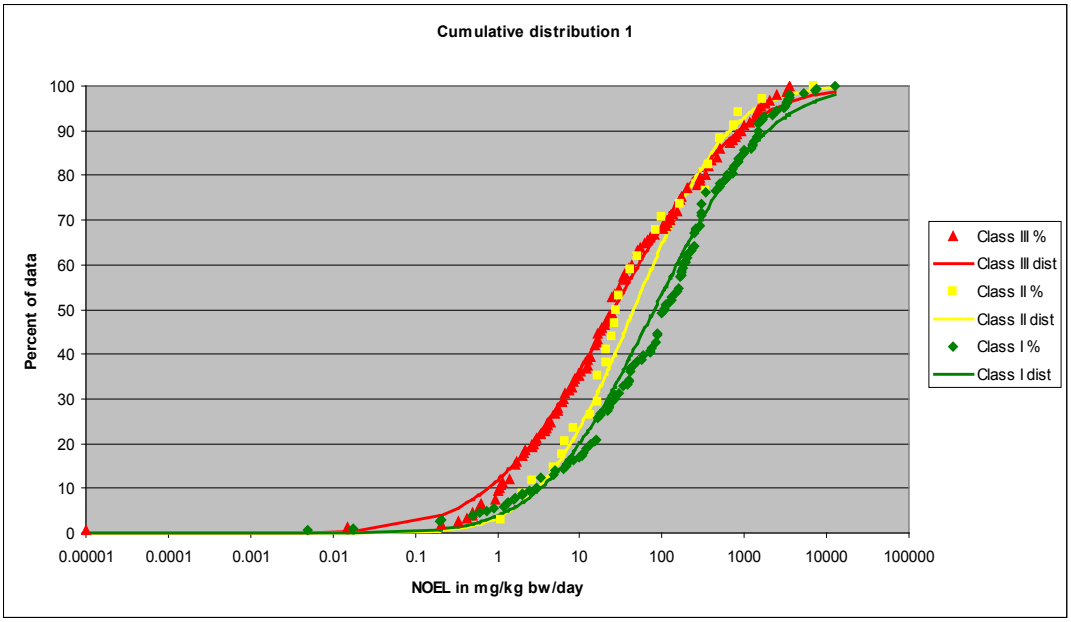


Figure 4.7. Cumulative distribution analysis of the COSMOS TTC dataset

Table 4.3. Distribution analysis of COSMOS datasets compared with the Munro thresholds

Cramer class	No. of chemicals	5th percentile NOEL (µg/kg/day)	Human Exposure Threshold (µg/person/day)	Munro 5th percentile NOEL (µg/kg/day)	Munro TTC value (µg/person/day)
Class I	201	1362	817	3000	1800
Class II	34	1443	866	910	540
Class III	150	284	170	150	90

4.5 Discussion

The full ban on animal testing for all hazardous properties of cosmetic ingredients and their final products requires the cosmetic industry to replace all *in vivo* toxicological tests with alternative testing methods – *in vitro* and *in silico*. Within the context of COSMOS, the application of a

[3] This is the unadjusted value, i.e. calculated before applying the safety factor of 100.

Threshold of Toxicological Concern (TTC), a probabilistic risk assessment tool, originally proposed for negligible exposure to chemicals present in food, was evaluated, as a tool for risk assessment of cosmetic ingredients.

The present study aimed to verify the applicability to cosmetics of currently available Munro thresholds based on the three Cramer classes. For this purpose a non-cancer TTC COSMOS dataset was constructed, based on chemicals extracted from the Cosmetic Inventory and toxicity values (NOELs) extracted from several databases (ToxRefDB, PAFA, RepDose, ILSI).

The TTC COSMOS dataset containing 385 unique chemical structures with NOEL values from repeat dose toxicity studies was used to generate the Cramer distribution of chemicals and analyse the cumulative distribution of NOEL values in order to derive 5th percentile NOEL values. The TTC COSMOS dataset was also analysed in terms of key physio-chemical descriptors representing molecular weight, solubility, partitioning and reactivity, in order to verify if the Munro and COSMOS TTC datasets overlap.

Chemicals from three datasets (Munro, TTC COSMOS, and Cosmetics Inventory) were categorised according to the Cramer classification. In a typical distribution pattern, most chemicals are categorised for Cramer Class III. However, in the case of cosmetic ingredients, an even distribution of chemicals between Class I and Class III was observed. Further, the analysis of the theoretical cumulative distribution revealed that the 5th percentile NOEL for the COSMOS TTC in case of Class I chemicals is twofold lower than the original Munro 5th percentile NOEL. Such results could be attributed partially to the Cramer classification not being able to discriminate efficiently between Class I and Class III chemicals, as several chemicals with lower NOEL values than the relative 5th percentile were identified in Class I. On the other hand, such results could potentially question the conservativeness of the Class I threshold and its applicability to cosmetic ingredients. In this case, it would be important to understand which chemicals are not represented in the original dataset and therefore should be excluded from the application of TTC.

In case of Class III compounds, the 5th percentile NOEL was above the original Munro 5th percentile NOEL therefore making it protective for this group of chemicals. These results also confirm the validity of the threshold for Cramer class III.

By the time of this final writing of the thesis the COSMOS Project has come to the end and final results of the applicability of the TTC approach to cosmetic ingredients were presented (Williams et al., 2016; Yang et al., 2016). Within the COSMOS Project the chemicals' related data have been further evaluated and toxicological studies for the derivation of relevant NOEL values have been extensively reviewed and a non-cancer database was curated. A database with high quality NOEL values was established based on which it was determined that the current thresholds related to Cramer class I and Cramer Class III are protective also for cosmetic ingredients (Yang et al., 2016). As exposure is an important aspect of the TTC, other work under the COSMOS considered route-to-route extrapolation and decision tree to estimate the systemic availability upon dermal exposure was established (Williams et al., 2016).

Final analysis, conducted by Yang et al (2016), of the protectiveness of the Munro thresholds indicated the Cramer Class II has a lower threshold for cosmetic ingredients and is even lower then Class III. This could possibly indicate that the Cramer classification is not able to distinguish Class II and Class III chemicals adequately. Alternatively it might indicate that current thresholds are not protective for Class II chemicals. It would be interesting to understand if the reclassification of chemicals that would presumably fall to Class III would significantly impact the calculated threshold for Class III. It is, however, not surprising that the derived threshold values are above the current thresholds, as cosmetic ingredients are not substances designed to be active in a sense of pesticides for example. This is also confirmed with the analysis of the TTC value where COSMOS non cancer dataset was added to Munro non-cancer dataset and new threshold values were derived for Cramer classes. These were in the range of the previous values but less conservative (Yang et al., 2016).

5 Threshold of Toxicological Concern for Biocides

5.1 Introduction

In this chapter, chemical biocides are considered. A chemical biocide is a chemical substance applied to kill or inhibit the growth of an organism. Most biocides are used in large quantities, and are thus produced in large volumes. A rough estimate is of 330,000 – 1 million tonnes of active substances per year (DG Environment, 2007). Their use covers a vast range of applications from industrial products and consumer goods. The most important application area, in quantitative terms, is industrial and public water treatment (Ceresana, 2012). There are several biocidal active substances on the market that act differently and sometimes they are combined in a biocidal product to increase the overall effectiveness of the product. For example, some of the components that are added to household products (for example cleaning products), such as surfactants or membrane permeabilisers, may increase the efficacy of biocides in killing microorganisms.

Many biocidal substances are recognised to be potentially hazardous to human health as for example the toxicity of some pest controlling agents (organophosphates) correlates with cholinesterase effects in the blood and brain, where that of disinfectants and preservatives (isothiasolinones, hydrogen peroxide), used also in cosmetics, correlates with irritating effects on the skin (Hahn, Schneider, Gartiser, Heger, & Mangelsdorf, 2010).

Due to biocides vast range of applications and their recognised hazard, it is important that biocides are properly regulated. Therefore, government regulations are put in place requiring toxicological tests to demonstrate that biocidal products are safe for consumers, workers and the environment under specific conditions of use. In the EU, the *Biocidal Products Regulation (BPR, Regulation (EU) 528/2012),* discussed in Chapter 1, aims at controlling the European biocidal market.

5.2 Aim of the Study

The TTC approach was not historically applied to the biocides, so this remains a big gap in knowledge and gives space for further development and refinement of the TTC approach.

The aim of this study was to extend TTC to support the risk assessment of biocides. The study developed the concept of TTC further as it verified if the Munro dataset is representative of a biocides dataset. Based on a limited dataset of NOEL values, collected from publicly available sources, it verified if Munro thresholds, for the three Cramer hazard classes would also be protective for biocides.

5.3 Materials and Methods

A list of 321 biocide active compounds (hereafter, biocides), without toxicity data and chemical structures, was provided by the former biocides group within the JRC's IHCP in 2011.

From the initial set of biocides, inorganic compounds (54), polymers (10), mixtures (17) and microorganisms (4), protein (1), mixtures of plant (coal) extracts (7) were removed and four poorly defined SMILES were not included. Almost all quaternary ammonium salts (26) were removed, apart from BKC (benzalkonium chloride) and DDAC (dialkyldimethylammonium chloride) as these generic names address a multitude of structurally-related but different substances, in alkyl chain lengths, alkyl chain distributions and anions (EC, 2006) resulting in 198 chemicals, which are presented in the Appendix 9.3. Salts were cleaned/neutralised (46) by removing the cationic part and the charge on the anion. 198 biocide chemical structures remained for the analysis. Chemical structures were stored as .sdf and .smi files for further analysis.

Separate .smi and .sdf files, including 10 quaternary ammonium salts were generated for the purpose of chemical space analysis relative to structural features analysis.

The reference Munro dataset containing 596 structures was available as .sdf file from the COSMOS Project, as of September 2011.

5.3.1 Structural Features Analysis with the Leadscope Tool

The Leadscope explorer v2.4 software, developed by Leadscope Inc. (http://www.leadscope.com/index.php) was used to explore datasets of chemical structures and related biological or toxicity data. Leadscope provides a number of ways to group a set of compounds, including chemical feature hierarchy with 27,000 named substructures. It thus provided a fast approach to become familiar with the compound collection. However it should be noted that the structural features represented in the structural analysis with Leadscope might not cover all structural features in a dataset, as they are a pre-defined selection by the software developer.

Generated .sdf data files for 198 biocides, 10 quaternary ammonium salts, Munro 560 and a subset of 77 biocides with toxicological data (biocides 77) were imported in the Leadscope software. Structural features analysis was run to compare the presence of structural features between datasets, thus allowing conclusions regarding representativeness of Munro dataset for biocides.

5.3.2 Analysis of Molecular Descriptors Calculated with Adriana.Code[4] Tool

Chemical space was analysed also in terms of molecular properties and Adriana.Code was applied to calculate selected molecular descriptors to facilitate the overlap in properties analysis.

Adriana.Code v 2.2., developed by Molecular Networks, calculates a series of molecular descriptors that are easy to use with data analysis software, for example, for similarity perception and modelling and the prediction of physical, chemical and biological properties of chemical compounds. Because of the sound geometric and physico-chemical basis of the calculated molecular descriptors, models developed with these descriptors allow a direct

[4] Adriana.Code was used in the study. The software module is no longer supported and has been replaced by CORINA Symphony.

interpretation of the effects influencing the analysed property. Adriana.Code also contains a series of methods for the generation of 3D structures and the calculation of physicochemical descriptors and molecular properties based on rapid empirical models (Molecular Networks, 2013).

The Lipinski rules to assess key physico-chemical properties related to the probability of good absorption were chosen. The following descriptors were calculated: M.I. Weight, H bond donors, H bond acceptors, Number of violations of Lipinski rule, XLogP, McGowan's molecular volume, Topological polar surface area (TPSA), Complexity, Solubility and Dipole .

Lipinski's Rule of Fives describes key structural features that are relevant for good physico-chemical properties of drug-like compounds. A violation of a rule might not account for a poorer absorption but the likelihood of poorer absorption is increased if more rules are broken and the extent to which they are exceeded.

Molecular Weight is related to the size of the molecule. As molecular size increases, solubility decreases, it also impedes passive diffusion through the bilayer membrane.

H bond donor is an atom connected to at least one hydrogen atom and an **H bond acceptor** is an atom with a lone-pair electron.

XLogP, the partition coefficient (P) is the ratio of concentrations of a compound in the two phases of a mixture of two immiscible solvents at equilibrium. Normally one of the solvents is water while the second is hydrophobic such as octanol. Hence the partition coefficient measures both how hydrophilic ("water-loving") or hydrophobic ("water-fearing") a chemical substance is. Hydrophobic chemicals with high partition coefficients are preferentially distributed to hydrophobic compartments such as lipid bilayers of cells while hydrophilic chemicals (low partition coefficients) preferentially are found in hydrophilic compartments such as blood and the cytoplasm.

McGowan's molecular volume determines transport characteristics of molecules, such as intestinal absorption or blood-brain barrier penetration. McGowan molecular volume in [ml/mol] is approximated by fragment contributions.

Topological polar surface area (TPSA) in [$Å^2$] of the molecule is derived from polar 2D fragments. The topological surface area of a molecule, defined as sum of the contributions to the molecular surface area of 2D polar atoms, such as oxygen, nitrogen and their attached hydrogens, is a measure of tendency for polar interactions (Prasanna & Doerksen, 2009).

Molecular polar surface area (PSA), i.e. surface belonging to polar atoms, is a descriptor that was shown to correlate well with passive molecular transport through membranes and, therefore, allows prediction of transport properties of drugs.

Complexity is another topological descriptor that provides a measure of skeletal complexity as a function of bond connectivities and the diversity of atom types (Hendrickson, 1987).

The calculated **solubility,** given in mol / L [log units], represents the maximum dissolved concentration in water. It is one of the most important molecular properties in the discovery of new active substances as low solubility limits absorption and causes low oral bioavailability.

Dipole moment, given in [Debyes], informs about the polarity of the molecule. Interactions between polar molecules create strong intermolecular forces therefore a number of physical properties are influenced (tension, solubility, melting and boiling points).

5.3.2.1 Statistical Analysis of the Calculated Descriptors

STATISTICA 10 was used to characterise the descriptors and to build 3D plots to visualise the space coverage and overlap of calculated properties. It was also used to perform the student-t test.

Data from excel files with calculated descriptors for the biocides and Munro dataset were imported to a STATISTICA default spreadsheet. Then STATISTICA workbooks were created which provided an easy manipulation of those data.

The Student t-test was run to identify those descriptors that differ significantly in terms of their mean values. The t-test is a statistical hypothesis test that tries to determine if the averages of the two samples are significantly different. The null hypothesis, that the difference is due to chance or that there is no difference is rejected, in cases where probability "p" value is "low", thus confirming that the means are in fact significantly different.

5.3.2.2 Data

.sdf data files for 198 biocides and 596 Munro's structures were imported to Adriana.Code and translated into 3D molecular structures. The import with translation resulted in a successful import of 181 biocides and 596 Munro's structures. Therefore, molecular descriptors were calculated for 181 biocides structures and Munro's 596 structures.

Descriptors for 596 Munro and 181 biocides (18 compounds could not be translated into 3D structures) were calculated. Data from excel files with calculated descriptors were imported into a STATISTICA default spreadsheet. Then STATISTICA workbooks were created to allow for easy manipulation of the data. The overlap in chemical space described by some molecular descriptors was analysed by visualisation using Statistica 10.

5.3.3 Cramer Classification

The Cramer classification scheme, implemented in the Toxtree software (v 2.1.0),[5] was originally applied to a dataset of 198 biocides and to a subset of 77 biocides, to group substances into hazard classes for the potential TTC estimation. The classifications were further re-checked by applying Toxtree software (v 2.6), the use of the updated version of ToxTree did not affect the classifications. A subset (77 structures from the original 198 structures) was determined based on the availability of repeat dose (NOEL) values in the Munro database, ToxRefDB,

[5] http://toxtree.sourceforge.net/

RepDose and COSMOS TTC DS. The number of NOEL values among sources is shown in Figure 5.1.

5.3.3.1 Data

RepDose, ToxRefDB, Munro and Non Cancer COSMOS TTC cosmetic ingredients datasets were searched for experimental data (NOEL or LOEL) that would allow the derivation of the NOEL for the purpose of this study. Full details for these databases are provided below. In all instances, the lowest NOEL (LOEL) value was taken for each compound.

RepDose (repeated dose toxicity) is a relational database on subacute to chronic toxicity studies and currently contains 655 chemicals. The toxicity of these chemicals is documented for about 2280 studies, carried out in rats, mice or dogs with oral or inhalation exposure. The database has been developed by Fraunhofer ITEM (Bitsch et al., 2006) with funding from Cefic LRI and grows continuously. It is accessible through the following web link http://www.fraunhofer-repdose.de/ upon registration. The database was searched as available in autumn 2011.

ToxRefDB (Toxicity Reference Database) is a database containing the results of animal toxicity studies. The database is a joint effort between the United States National Centre for Computational Toxicology (NCCT) and Environmental Protection Agency's (US EPA) Office of Pesticide Programs (OPP). ToxRefDB includes NOEL values for rat chronic and cancer assays, mouse cancer assays, rat 90-day toxicity studies, rat multigenerational reproduction studies, and rodent and rabbit prenatal developmental studies. For the purpose of this study, data from an excel file associated with a publication from Knudsen et al. (Knudsen et al., 2009) were analysed. The file for automated extraction of data contains 386 tested pesticide actives by name (386 unique CAS RNs) with Lowest Effect Level (LEL) values associated with rat prenatal developmental/maternal toxicity studies.

The Munro dataset is currently the *de facto* TTC database for non-cancer endpoints. The dataset, contains data for 613 tested chemicals. The structure data file and summary data tables are downloadable from EFSA (http://www.efsa.europa.eu/de/supporting/pub/159e.htm). The

summarised data include study design, NOEL values and the associated critical effects. These files were compiled for the purposes of a chemoinformatics investigation carried out by Soluzioni Informatiche srl under the terms of an EFSA contract.

COSMOS TTC Dataset (Version 1.0) consisted of 660 compounds used as cosmetic ingredients, among which 385 compounds were identified with a well-defined structure and quality checked NOEL values. As the COSMOS Project was at that time of performing this analysis an ongoing project, the dataset was subject to changes. For the purpose of the study presented here, the COSMOS TTC dataset as available in 2011 was used. Additional information relative to the dataset can be found in Chapter 6.

5.3.4 Derivation of Human Exposure Threshold Levels for Biocides

For the purpose of the validation of the relevance of TTC concept, specifically the applicability of threshold values for Cramer classes to biocides, a subset of chemical biocides with selected NOEL values was analysed further. In this case, it has to be noted that due to limited availability of toxicological experimental data the NOEL/LOEL values as available from public sources were extracted, which means that the value does not necessarily represent the NOAEL, which could be lower or higher. This means that the derived 5th percentile NOELs might be more or less conservative.

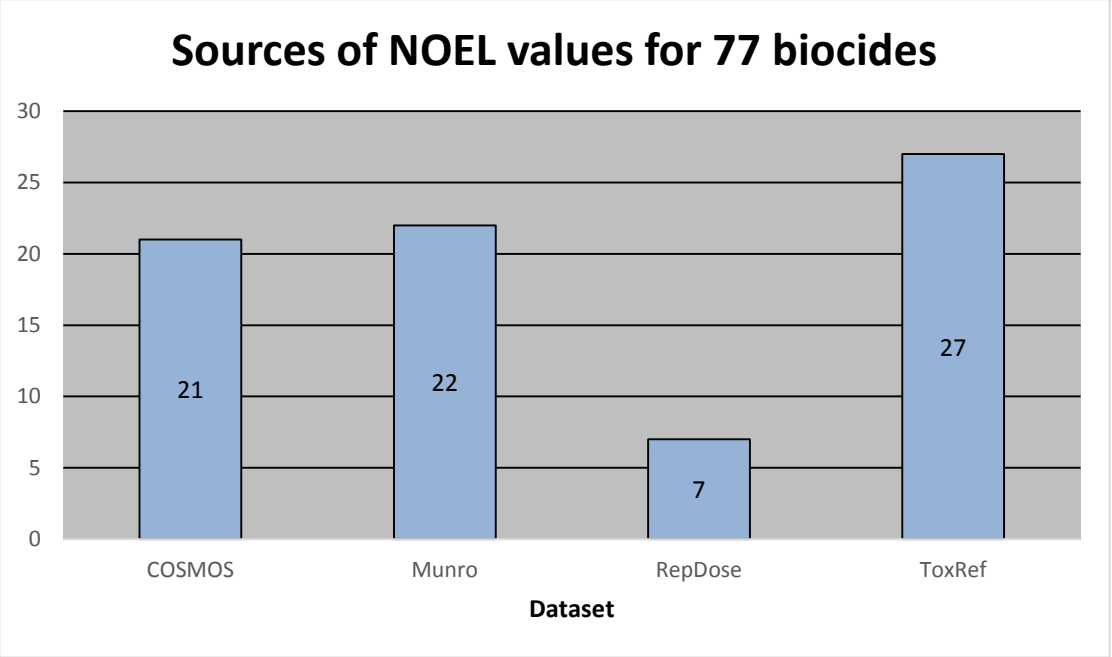


Figure 5.1. Sources of NOEL values for 77 biocides

5.3.4.1 Standardisation of NOEL and LOEL Values

The TTC concept builds thresholds values on chronic oral repeat dose studies. Therefore, various sources were searched for high-quality oral, rat, toxicity studies that describe a variety of toxicity endpoints. Among the NOELs, those derived from chronic studies were preferred. In cases were no such study was available a LOEL or a NOEL from a sub-chronic study was preferred.

For substances extracted from ToxRefDB where only LEL values were available, an extrapolation factor of 3 was applied to derive a NOEL (LEL/3). In the RepDose, sub-chronic and chronic studies were preferred. In cases where only a sub-chronic NOEL was available, an extrapolation factor of 3 was applied and in cases where a sub-chronic LOEL was available, first an extrapolation factor of 3 was applied to derive a chronic LOEL and then again a factor of 3 to derive a chronic NOEL. A factor of 3 was also applied to NOELs derived from sub-chronic studies in the Munro dataset.

The resulting TTC dataset for biocides thus contains 77 compounds with *in vivo* studies of which 27 from ToxRefDB, 7 from RepDose, 22 from Munro and 21 from the previous studies on

cosmetics. In cases where more than a single study was available for a compound, only the most conservative study was considered in order to cover the worst-case scenario. 27 NOEL values were derived from LEL values, 4 were derived from LOEL sub-chronic, 4 from sub-chronic NOEL values the remaining 42 NOEL values were taken just as provided.

5.3.4.2 How the TTC was Derived

Visualisation of the NOEL sources, Cramer classification and statistical analysis was performed using the Statistica 10, a software programme from StatSoft, and Excel from Microsoft.

The log normal cumulative distribution functions (CDF) for Classes I and III were plotted. The theoretical CDFs were fitted to the empirical data. The CDF, or just distribution function, was selected as it was originally used by Munro (1996) and it describes the probability that a real-valued random variable X with a given probability distribution will be found at a value less than or equal to x. The 5th percentile NOELs were derived from log normal theoretical CDF distributions by considering the average and standard deviation of the empirical distribution. The 5th NOEL served as a point of departure to derive a human exposure threshold, which was calculated by applying a standard uncertainty safety factor of 100 for extrapolation from animals to humans (10x) and for the consideration of interspecies differences (10x).

5.4 Results

5.4.1 Structural Features Analysis with Leadscope

In the Leadscope analysis, an additional file containing 10 quaternary ammonium salts was selected in parallel to the original 198 biocides to represent the importance of quaternary ammonium salts among biocides.

The use of structural features analysis (Figure 5.2) showed that in the Munro dataset quaternary ammonium salts are underrepresented (notably quaternary ammonium cations) among surfactants. Several structural classes are missing, including oxidising biocides (peroxides),

heterocycles, sulphur (thio) containing compounds (isothiazole, isothiazolidine), carbamates (dithiocarbamic acid) and halides (some bromides).

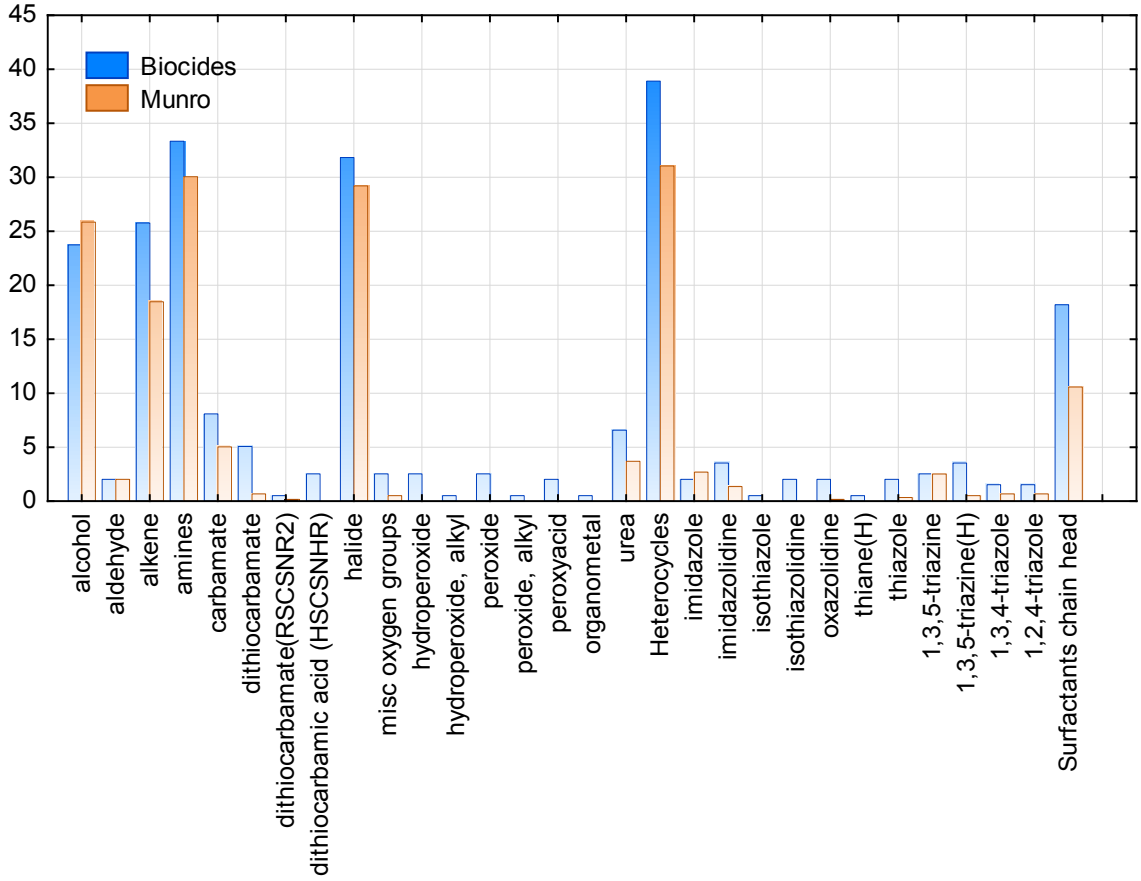


Figure 5.2. Identified structural features and their relative distribution in the datasets presented as structural features (x axis) in relative values (y axis)

5.4.2 Analysis of Molecular Descriptors with Adriana.Code Tool

The biocides and Munro chemical space were also analysed in terms of the overlap of physiochemical properties, described by selected key descriptors; namely partition coefficient XLogP, McGowan molecular volume, Topological Polar Surface Area (TPSA), Complexity, Solubility and Dipole. These molecular descriptors represent some general structural properties, relevant for the chemical to be effective against the target, including bioavailability and reactivity. Calculated descriptors for the two datasets were plotted in 3D space to analyse the overlap of their distribution. These are shown in Figure 5.3 and Figure 5.4.

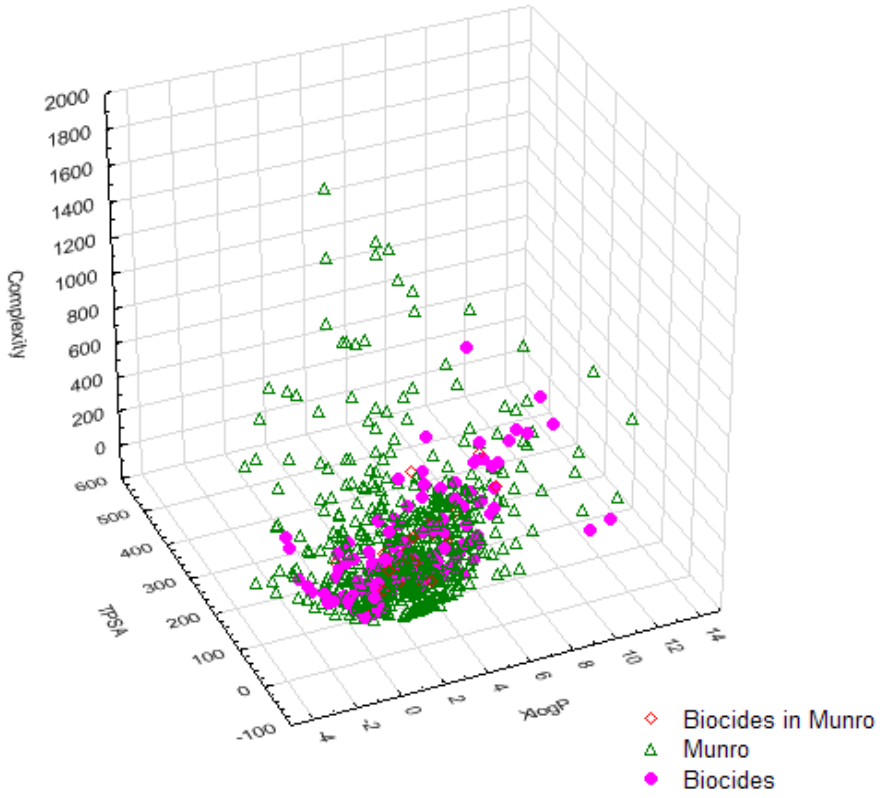


Figure 5.3. 3D scatterplot complexity against XlogP and TPSA for the Biocides and Munro datasets

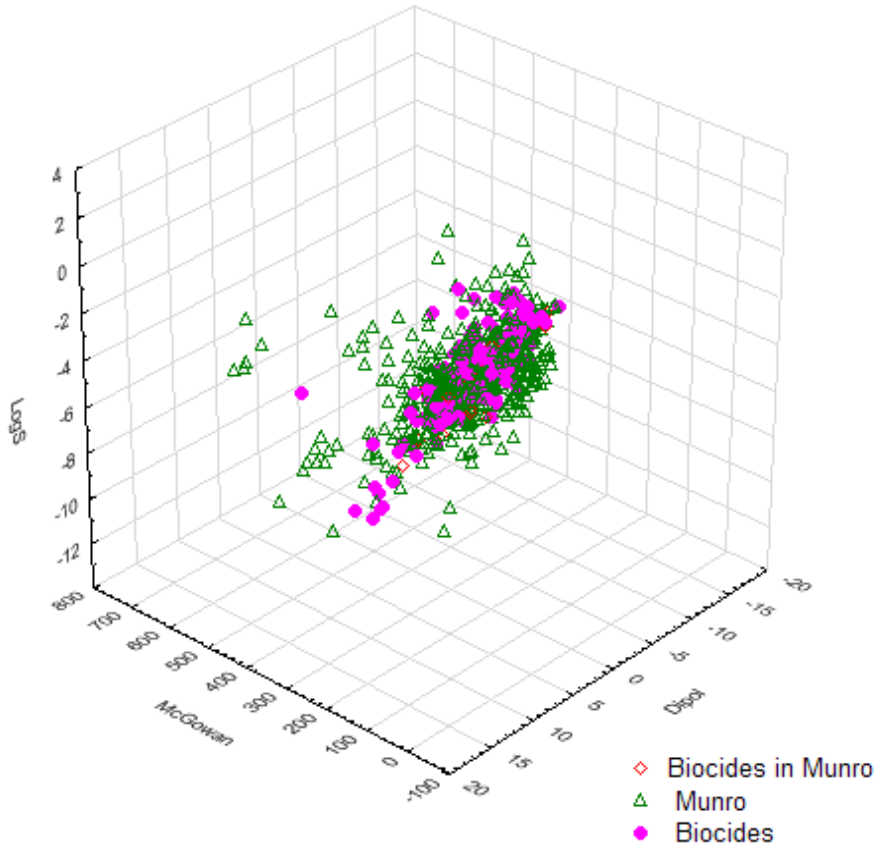


Figure 5.4. 3D scatterplot LogS against McGowan and Dipole for Biocides and Munro dataset

The biocides dataset covers a diverse range of physio-chemical properties. Compared with the space occupied by Munro inventory, the physio-chemical properties describing the biocides dataset cover a narrower range of values that is also confirmed by smaller standard deviations values, shown in Table 5.1, for the mean of each descriptor. It can be concluded that the Munro and biocides datasets overlap in the space described by XLogP, McGowan molecular volume, Topological Polar Surface Area (TPSA), Complexity, Solubility and Dipole. Thus the Munro dataset includes the biocide dataset.

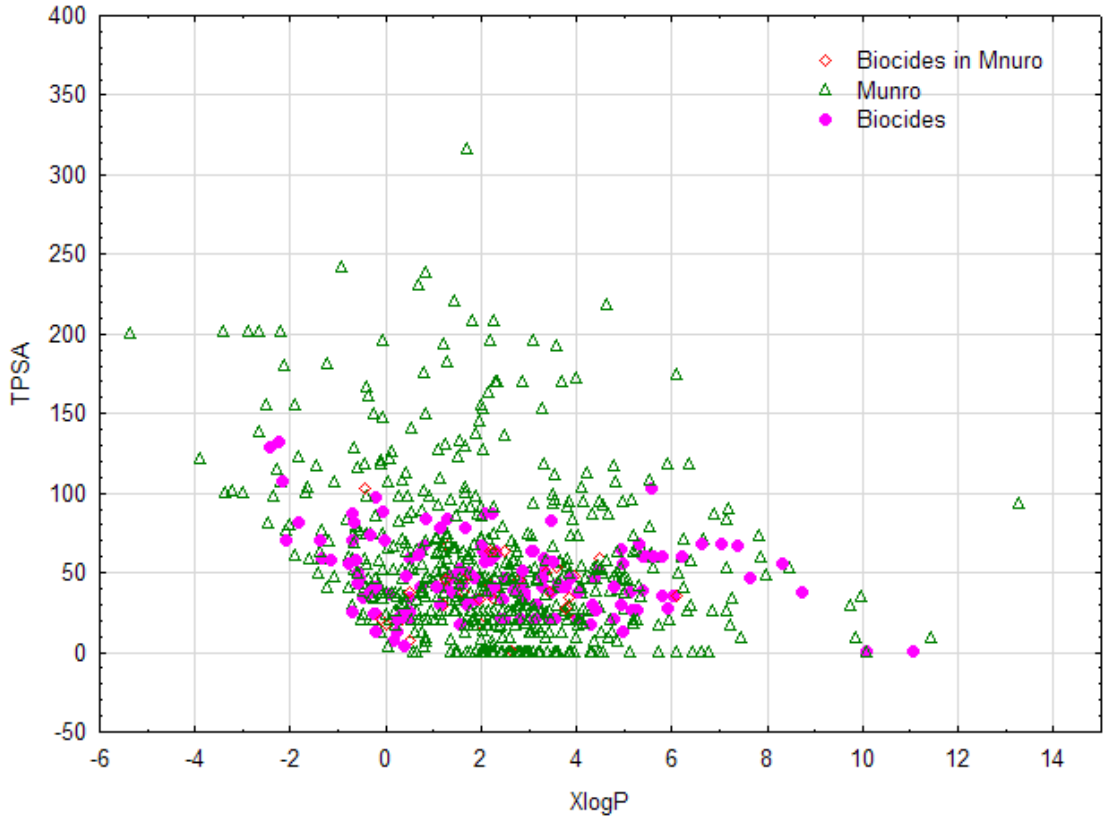


Figure 5.5. Graphical representation of the distribution of physicochemical properties of Biocides and Munro dataset

Calculated descriptors were further analysed for significant differences in their respective mean values by applying the student t test. Among the molecular descriptors compared, topological descriptor showed a significant difference, this was the Topological Polar Surface Area, as shown in Figure 5.5. This difference could suggest that biocides have higher bioavailability as they may be able to penetrate the cell membrane easier, described by lower TPSA values.

Table 5.1. Calculated physicochemical properties and their respective averages, standard deviation and the t-test.

	Mean MUNRO	Mean BIOCIDES	t-value	p	Std.Dev. MUNRO	Std.Dev. BIOCIDES
Weight	251.5713	236.5321	1.2405	0.2151	150.1718	115.3595
HDon	1.3272	0.9171	2.8439	0.0046	1.8431	1.0948
HAcc	3.8574	3.2541	2.2899	0.0223	3.4036	1.7862
XlogP	2.2424	2.3750	-0.6519	0.5147	2.3934	2.4087
TPSA	56.6025	44.6980	2.9605	0.0032	52.4787	23.6999

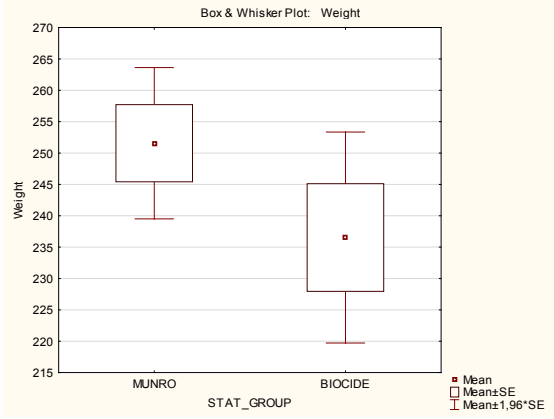
	Mean MUNRO	Mean BIOCIDES	t-value	p	Std.Dev. MUNRO	Std.Dev. BIOCIDES
McGowan	179.3742	170.3073	1.0773	0.2817	103.3232	83.9982
Dipole	3.7566	3.9530	-0.4499	0.6529	5.4825	3.8111
LogS	-2.8730	-2.8177	-0.2905	0.7715	2.2174	2.3239
NViolationsRo5	0.2483	0.1657	1.740215	0.082218	0.5902	0.4411
Complexity	288.7956	261.4555	1.2236	0.2215	278.0488	207.0840

Table 5.2. Descriptive statistics comparison for TPSA (Munro vs. Biocides dataset)

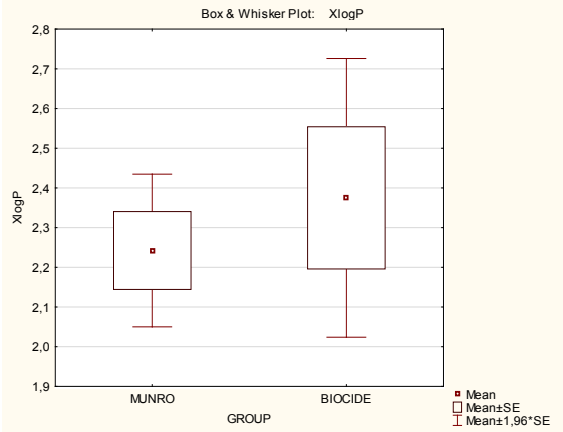
Descriptive statistics value	Dataset	
	MUNRO	BIOCIDE
Valid N	596	181
Mean	56.60	44.70
Confidence (-95,000%)	52.38	41.22
Confidence (95,000%)	60.82	48.17
Median	46.21	40.62
Minimum	0.00	0.00
Maximum	554.05	132.13
Lower (Quartile)	23.79	27.05
Upper (Quartile)	71.90	59.32
Std.Dev.	52.48	23.70
Confidence SD (-95,000%)	49.66	21.48
Confidence SD (+95,000%)	55.64	26.43
Standard (Error)	2.15	1.76

By looking into the means and standard deviation of each dataset, it was shown that the biocides dataset contains smaller structures in terms of molecular weight. The partitioning coefficient for biocides is within the optimal range (0-3) for partitioning; however, the mean value itself is

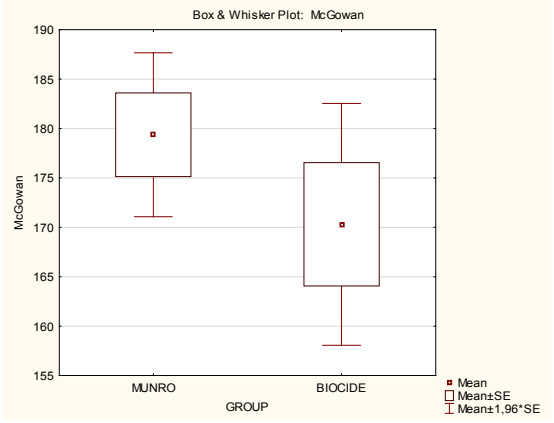
slightly higher, indicating the importance of lipid bilayer permeability by the prevalence of lipophilic chemicals (surfactants) (higher log P). Statistically significant difference was confirmed by the student t-test (Table 5.1) for topological polar surface area and H bond donors and acceptors as shown in Figure 5.6. This is not surprising as major contributors to the polar surface area are hydrogen bond donors or acceptors. Biocides have fewer violations of the Lipinski rule of 5 demonstrating a higher number of more favourable structures for oral absorption. This is further confirmed by statistically significant lower number of H bond acceptors and H bond donors among biocides. Smaller standard deviations for each descriptor could be used to confirm observations of 2D and 3D descriptors plots, that is, molecular descriptors representing the biocides dataset are less spread out, thus occupying a narrower chemical space.



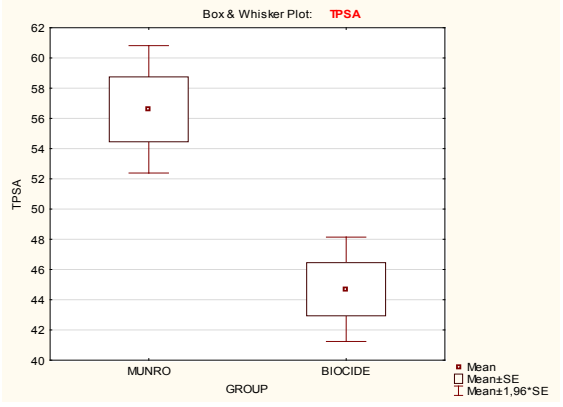
Box & Whisker Plot: Weight



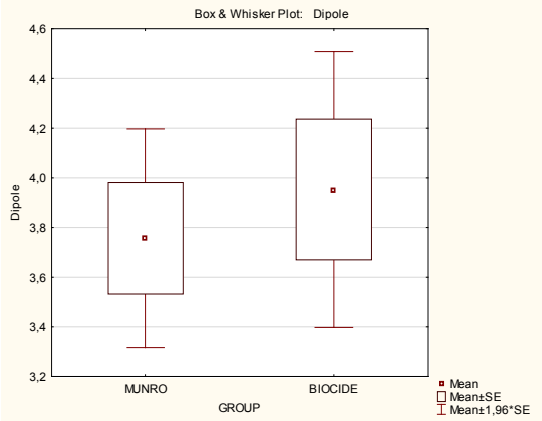
Box & Whisker Plot: XlogP



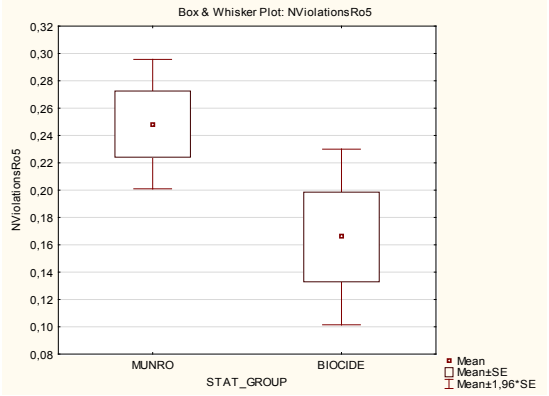
Box & Whisker Plot: McGowan



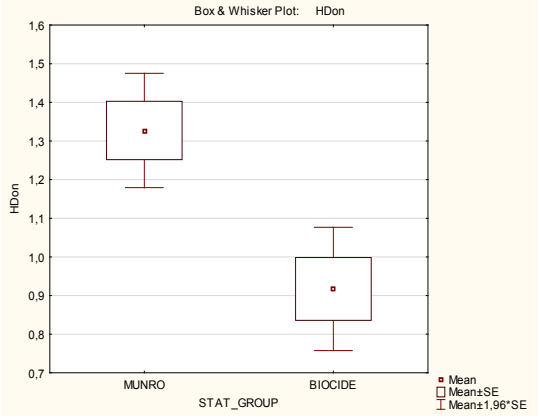
Box & Whisker Plot: TPSA



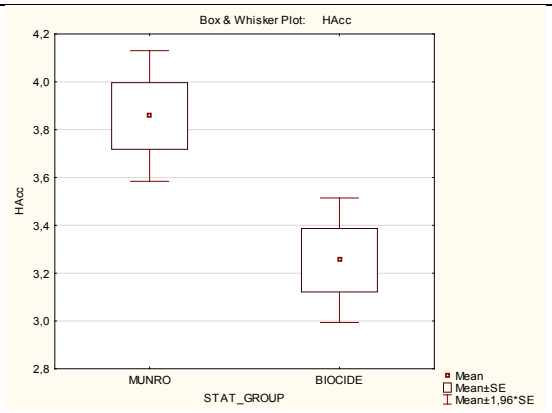
Box & Whisker Plot: Dipole



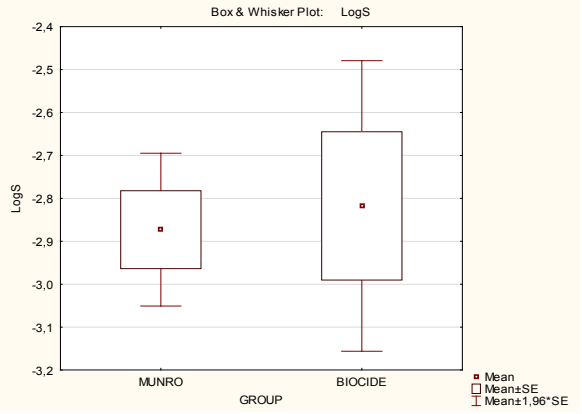
Box & Whisker Plot: NViolationsRo5



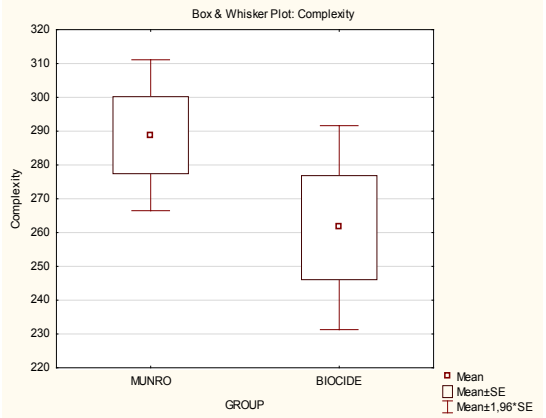
Box & Whisker Plot: HDon



Box & Whisker Plot: HAcc



Box & Whisker Plot: LogS



Box & Whisker Plot: Complexity

Figure 5.6. Box & Whisker Plot for the calculated physicochemical properties of Biocides and Munro dataset

Based on the above observations, it could be concluded that Munro and biocides dataset broadly overlap in terms of physio-chemical properties, however the physio-chemical properties

describing the biocides dataset have a narrower range of values. The significant difference in means for TPSA is further supported with the descriptive statistics in the Table 5.2.

5.4.3 Cramer Classification Results

198 biocides and 77 biocides compounds with NOEL values were classified with the Cramer decision tree.

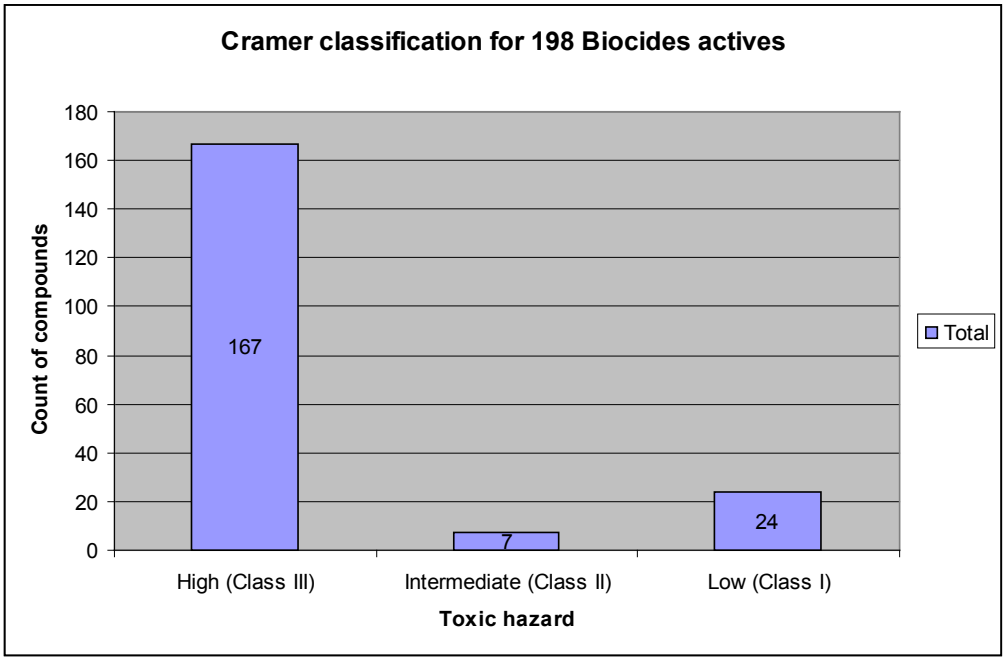


Figure 5.7. Cramer classification for 198 biocides actives

The Cramer classification was applied to the 198 biocide active compounds. 84% of compounds were allocated to Class III; 12% to Class I and 4% to Class II. As expected, most of biocides active compounds were classified as Class III (High) for toxic hazard. The results of the classification into Cramer classes are shown in Figure 5.7.

Then 77 biocides were also classified with the Cramer decision tree for 5th percentile NOEL calculation in the next step. The results of the classification are shown in Figure 5.8. The majority, 78% of compounds, belong to Class III (High), 18% to Class I (Low) and 4% in Class II (Intermediate). Again, the highest percentage of compounds belongs to Class III. There are slightly more compounds in Class I for this group compared to the original biocides dataset.

Both distributions, of the whole biocides dataset 198 and its subset 77, reflect well the distribution of the Munro's dataset of which 73% were placed in Class III, 4% in Class II and 22% in Class I.

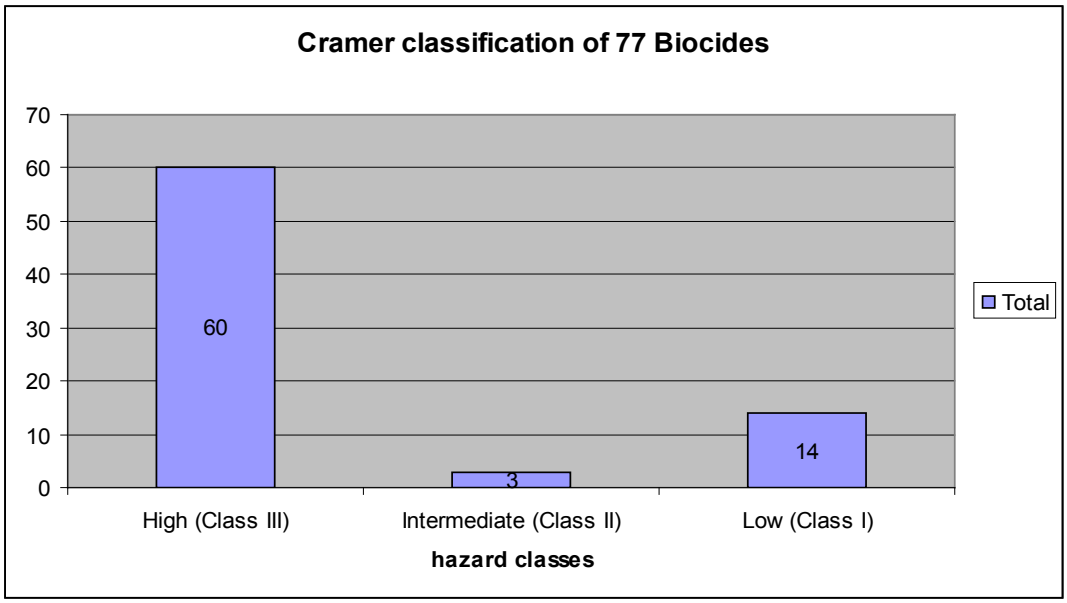


Figure 5.8. Cramer classification of 77 biocides

The distribution in the Figure 5.9 shows how the compounds from different sources are classified for the level of toxic hazard. According to the Cramer decision tree, most values extracted from ToxRefDB database, 35 were allocated to Cramer class III (High) for toxic hazard. Among 12 compounds extracted from RepDose, 6 compounds belong to Cramer class I and 5 are in Class III. Compounds extracted from Munro dataset show the same distribution as those from ToxRefDB, which means 19 out of 25 are in Class III. Among 4 compounds from the non-cancer COSMOS TTC dataset, 3 compounds are in Class I.

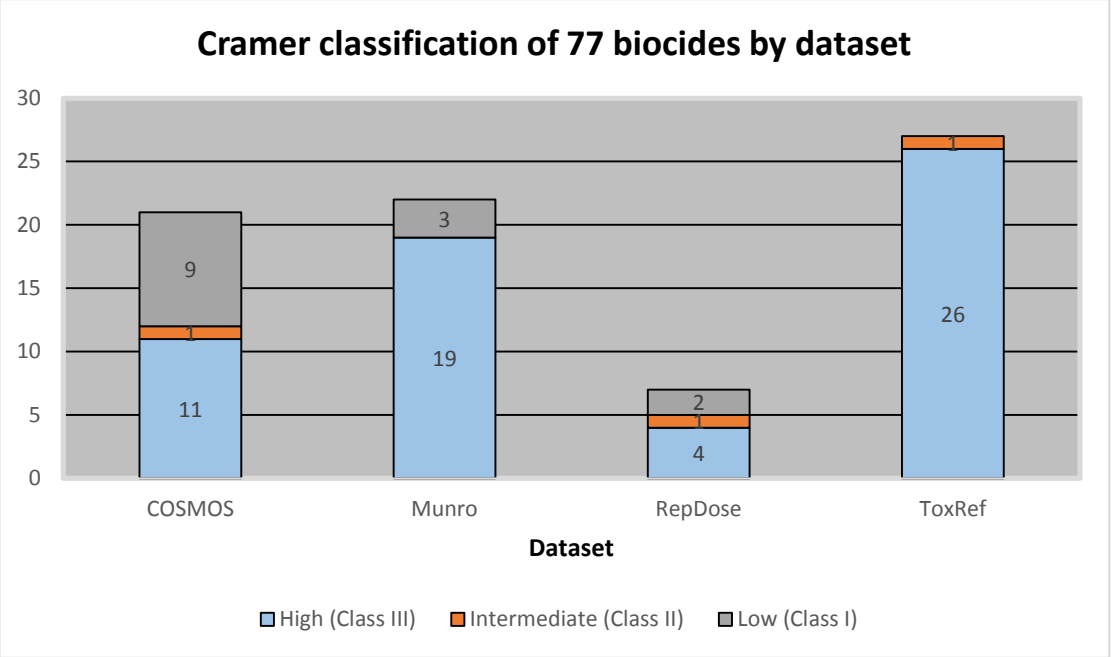


Figure 5.9. Cramer classification of 77 biocides according to the database from which they were derived.

5.4.4 Human Exposure Threshold Levels for Biocides - Results

The estimation of structure related 5th percentile NOEL for biocides was performed in a way that is consistent with the analysis performed by Munro et al. (1996) and included 14 compounds classified in Cramer Class I and 60 compounds classified in Class III. Cramer class II was not considered due the low number (3) of compounds. No compound was rejected due to presence of specific functional groups (genotoxic alerts) or because it is an organophosphate or carbamate. The presence of neurotoxic chemicals is relevant to keep Class III threshold values conservative for the protection from other neurotoxic effects (Feigenbaum, Pinalli, Giannetto, & Barlow, 2015), and no less important to keep the results conservative in general.

The cumulative distribution functions of Class III and Class I NOEL values were plotted as shown in Figure 5.10. The NOEL distribution plotted in a log normal form is not well separated in the case of low NOEL values but instead they intersect each other. This is an indication that the Cramer decision tree fails to distinguish well between Class I and Class III compounds if it is assumed that Class I compounds should have higher NOEL values then Class III compounds.

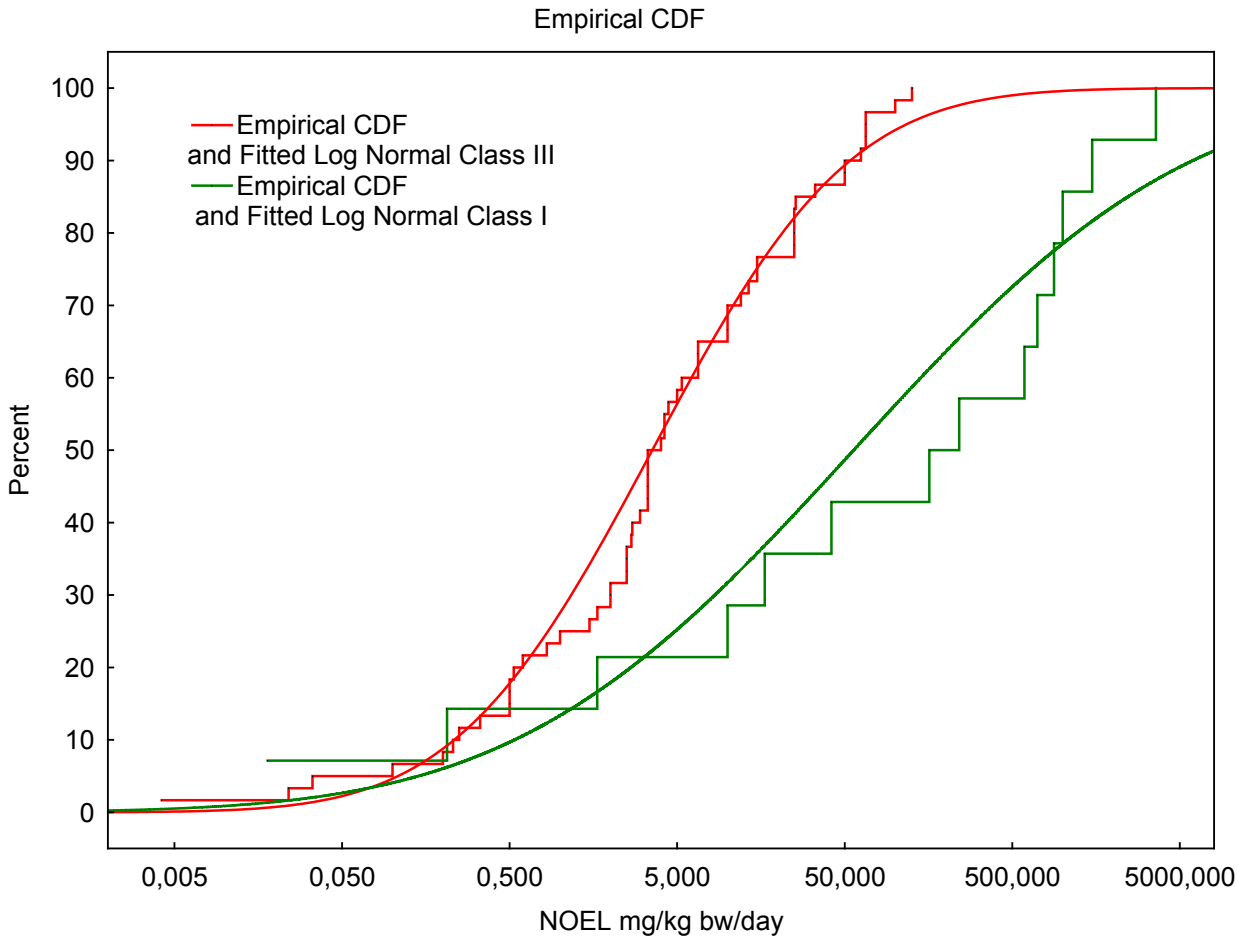


Figure 5.10. Cumulative distribution functions of Class III and Class I NOEL values

Then the theoretical log normal distribution function was fitted to the empirical function. By taking the mean and the standard deviation of the theoretical CDF function, the 5th percentile NOEL for each class was derived. The results indicate that they are lower than Munro's classifications.

The TTC values from the datasets of 14 and 60 compounds were calculated using the respective 5th percentile NOELs and the same assumptions used by Munro et al (1996) which include an uncertainty factor of 100x (10x for interspecies differences and 10x for human variation) and a body weight of 60 kg. Calculated human threshold values for the two Classes are also lower than the Munro's threshold for the specific classes, currently in use when applying TTC.

Table 5.3. Derived 5th percentile NOELs and Human Exposure threshold for biocides

Cramer Classes and No of compounds	5th percentile NOELs (Biocides) (µg/kg/day)	Human Exposure threshold level (Biocides) (µg/person/day)	Human Exposure threshold level (Munro, 1996) µg/person/day
Class I (Low) 14	142	85	1800
Class III (High) 60	107	64	90

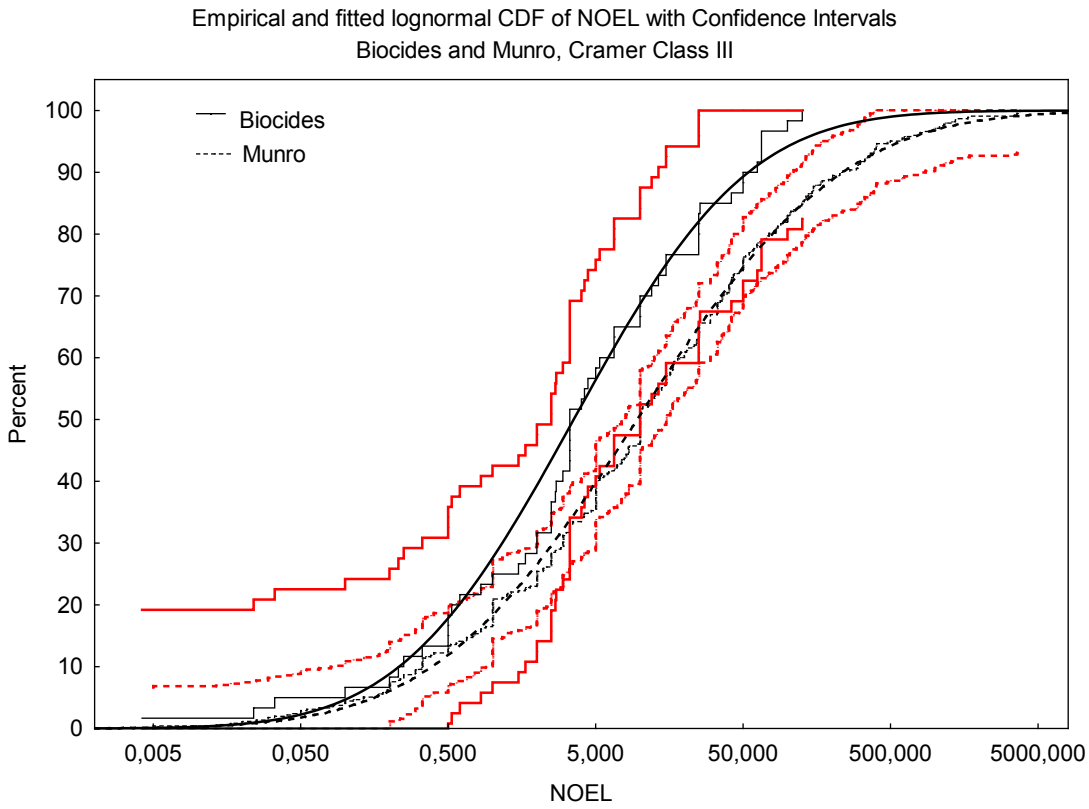


Figure 5.11. Empirical and fitted lognormal CDF of NOEL with confidence intervals for the Biocides and Munro dataset

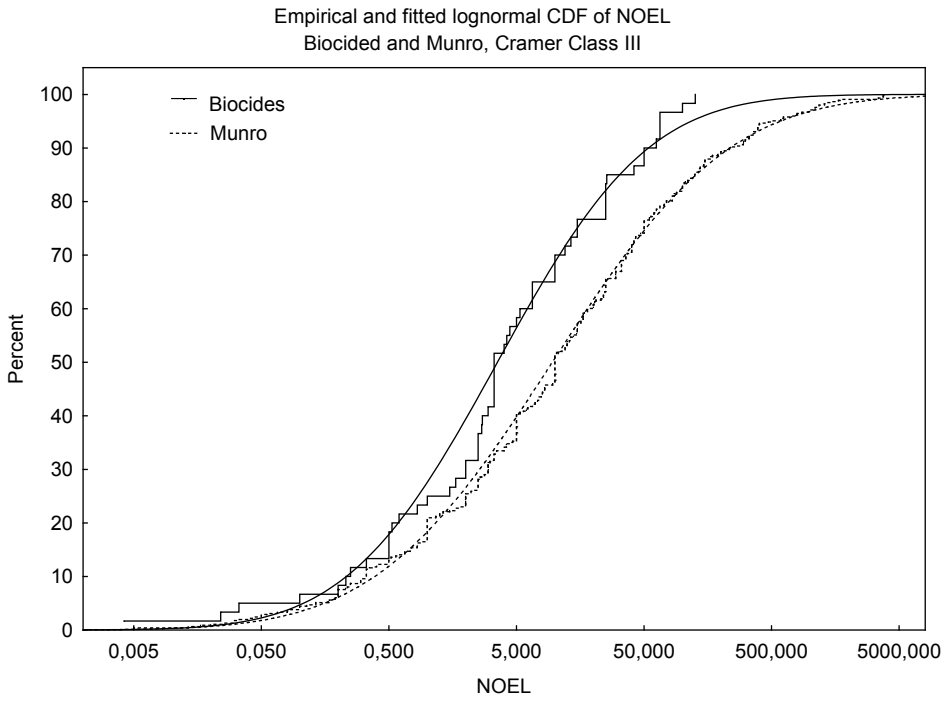


Figure 5.12. Empirical and fitted lognormal CDF of NOEL values for the Biocides and Munro dataset

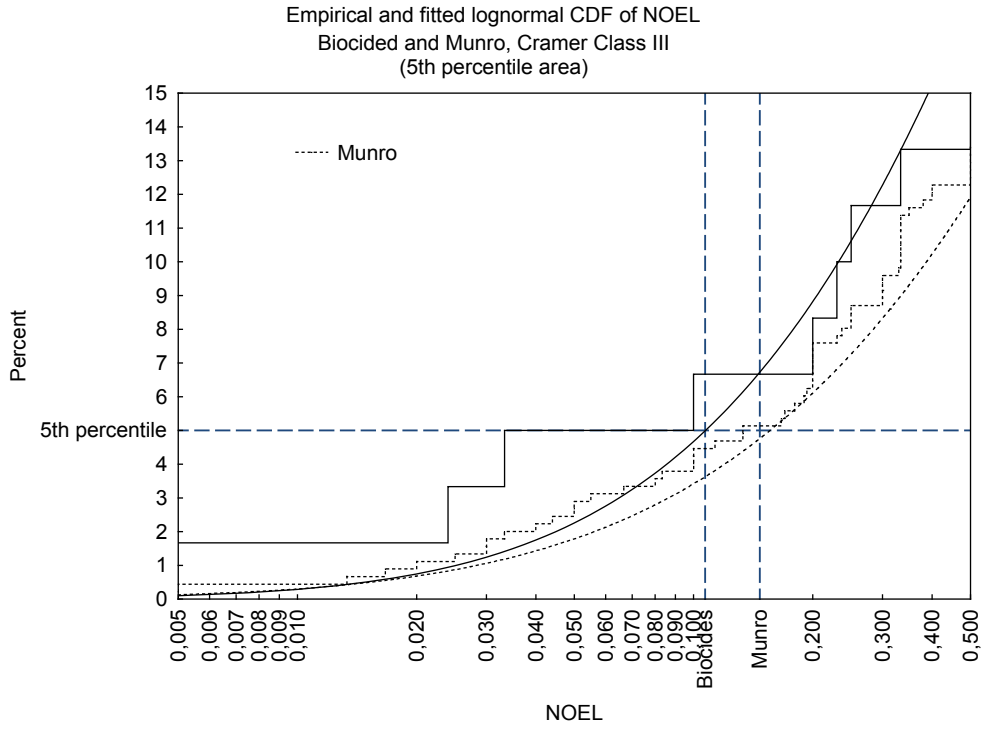


Figure 5.13. Empirical and fitted lognormal CDF of NOEL values for the Biocides and Munro dataset, the detail around the 5th percentile

By comparing the calculated exposure thresholds for biocides with Munro's, it is observed that the difference for Class III is not as significant as for Class I, 26 % and 95% respectively (Table 5.3). In addition, the standard deviation is bigger for data in Class I, which indicates data are more spread out. In the case of Class I the range of values is between (0.01 – 4000 mg/kg bw/d) which is driving a low 5th percentile NOEL. The drawback in this calculation is also the fact that there are too few data points to robustly determine the 5th percentile NOEL. Therefore this estimate is not considered reliable.

For Class III compounds a further assessment was performed to understand the level of difference in the 5th percentile value for Munro and Biocides distributions. It was confirmed, that the biocides 5th percentile lays in the confidence interval of the Munro's empirical distribution as shown in Figure 5.11, therefore confirming that there is no significant difference in values. However, it is observed in Figure 5.12 and Figure 5.13 that the theoretical log normal distribution for biocides is placed on the left of the to the distribution curve for Munro dataset, potentially showing that in general toxicological values for biocides are lower.

5.5 Discussion

In the light of decreasing *in vivo* testing, the Threshold of Toxicological Concern (TTC) approach has been proposed in the area of risk assessment. The BPR (EU, 2012) gives way to apply TTC under the Exposure Based Waiving approach. Exposure Based Waiving is proposed as an option in order to reduce animal testing where relevant exposure to humans remains within acceptable limits.

This preliminary study was undertaken to determine the relevance of available threshold values for the risk assessment of biocides substances. Several chemoinformatics methods have been applied to a dataset of biocides compounds to validate the representativeness of the chemical space described by the Munro dataset in terms of structural features and physio-chemical properties. Further, for a subset of biocides compounds, repeat dose toxicity data (NOEL values) were gathered from publicly available sources that allowed the derivation of the 5th percentile NOEL upon which threshold limit values for biocides chemicals belonging to two Cramer

hazard classes, namely Class I and Class III were derived. These values could be potentially used to reduce animal testing under Exposure Based Waiving principle.

Firstly, Munro and biocides chemicals datasets were compared in terms of Leadscope structural features describing the two chemical spaces. It was discovered that the chemical space described by the Munro dataset is includes the space described by the biocides dataset. However, a gap was identified as datasets differ for some groups of chemicals. In the Munro dataset surfactants, quaternary ammonium salts (notably quaternary ammonium cations), and also oxidising biocides (peroxides) are missing. Among heterocycles, potent organic biocides - isothiazole-containing compounds, isothiazolidine and among carbamates – dithiocarbamic acid and among halides some bromide-compounds are also lacking. (This should be considered to understand the impact of the NOELs for the above biocides actives on the 5th percentile NOEL).

Further, Munro and biocides chemical spaces, described in terms of selected physio-chemical properties, representing size (McGowan molecular volume), partitioning behaviour (XLogP), solubility, surface belonging to polar atoms (Topological Polar Surface Area (TPSA)), shape (Complexity), and polarity (Dipole), have been compared. The comparison demonstrated that physio-chemical properties describing the biocides dataset cover a narrower range of values, which was also observed by comparing means and their standard deviation. The significant difference in means for TPSA, a structural feature, which has been associated with membrane permeability (Kerns & Di, 2008), could be related to some specific properties for biocides.

In the next step, Cramer classification for three hazard classes has been applied to biocides dataset. Most of chemicals were classified for high (Class III) toxic hazard, which resembles the distribution of Munro dataset well, although a slightly higher prevalence of Class III biocide chemicals is observed.

Considering the Cramer analysis of the biocides TTC dataset in the next step, the cumulative distribution analysis was performed and further the 5th percentile NOEL was derived for Cramer class I and III. This NOEL values served as point of departure to which uncertainty factors have

been applied in order to derive a threshold level where exposure is still considered negligible. For Class I a threshold value, which is well below the currently existing Munro value, was derived. However, this estimate is considered unreliable due to the small size of the Class I subset.

For Class III compounds the calculated 5th percentile NOEL is 30% lower compared to the original Munro's 5th percentile NOEL, thus making the threshold for the Class III biocides compounds below the currently applicable Munro's threshold. This is not a surprising result as a subset of biocides active compounds, expected to exert a toxic effect, was analysed.

In the context of the relevance of Munro threshold currently in use for application to biocides, it cannot be concluded, at this stage, that the thresholds are not protective. However, the study indicates that further research should be carried out to identify those chemicals that are currently not covered by the representative Munro dataset. It also gives an indication that a lower threshold should be considered, however this should be further confirmed by compiling larger NOEL dataset from toxicological studies for biocides.

5.5.1 Conclusion

The presented study is a preliminary study aimed to assess the validity of currently applied threshold values for biocides active substances. It was not intended to set new threshold values for biocides, instead it initiated research on the possible application of TTC to biocides, and demonstrated under which circumstances this non-testing method could be applied.

At this point it should be emphasised what was revealed by the analysis of the chemical space represented by biocides and compared to the reference Munro dataset. Not all groups of biocides are well represented in the underlying Munro dataset and that the biocides dataset differs in some physio-chemical properties such as bioavailability. This could be further explored in order to identify groups of chemicals to which the TTC approach could not be applied and potentially refine the thresholds.

Calculated threshold values for Class I were below the values currently in use but this result is unreliable due the small size of the dataset. For Class II only two compounds were identified, therefore no further calculations were performed for this class. Instead, the threshold established for Cramer Class III is below but still in the range of the current threshold, thus supporting the validity of the 90 μg/person/day exposure threshold limit.

These findings give motivation for further research and show the necessity and relevance of open access data to build a bigger dataset in order to give more robustness to the analysis. One such source of data would be the registration reports prepared under regulatory requirements of BPR, for the registration of biocides.

Toxicity data for a subset of biocides active compounds were gathered from public sources and therefore are very limited. The NOEL values gathered might not necessarily represent the regulatory established lowest NOAELs for a chemical, therefore giving additional conservativism to the results. In addition, no further quality evaluation of the data could be performed at this stage. Following the study is considered conservative and no additional uncertainty factors were applied due to a small dataset, when calculating the thresholds for human exposure.

Even by considering all the eventual drawbacks of the study, the final threshold values especially for Class III are in the range of the current thresholds and further confirms the validity for Cramer Class III.

6 Application of the TTC Aproach to Developmental and Reproductive Toxicants

6.1 Introduction

6.1.1 Reproductive Toxicity

Reproductive toxicity refers to the adverse effects a chemical exerts on the reproductive ability of an organism and the development of its offspring. Reproductive adverse effects are considered those affecting sexual function and fertility, where developmental effects are considered those induced during pregnancy or prenatal exposure. A battery of standard toxicity tests is usually required by regulatory frameworks and include one-generation reproductive toxicity study - OECD technical guidance 415 (OECD, 1983), extended one-generation repro-ductive toxicity study OECD technical guidance 443 (OECD, 2011), two-generation reproductive toxicity study OECD technical guidance 416 (OECD, 2001a) and prenatal developmental toxicity study OECD technical guidance 414 (OECD, 2001b). However, regulatory frameworks also offer the possibility to apply alternative methods to assess the toxicological profile of the compound using exposure criteria, thus, if there is negligible exposure. In such cases, TTC seems a rational option.

Due the severity of adverse effects, reproductive toxicity is considered as equivalent concern as carcinogenicity and genotoxicity. Therefore, for TTC and risk assessment purposes, a key question is whether is it possible to identify a low dose of exposure, a threshold value, below which no adverse effects related to reproductive toxicity are expected. In case this would be lower than the Munro threshold for Cramer Class III, it would need to use a separate threshold, similar to those for genotoxic carcinogens. This issue has gained a lot of attention since the awareness that chemicals may perturb hormonal homeostasis, therefore affecting reproductive health. Endocrine Disruption (ED), an endpoint of high regulatory concern, is considered as perturbation of hormonal homeostasis and in practice mainly refers to the reproductive (oestrogen and androgen signalling processes that regulate the development and functioning of

the reproductive organs) and thyroid systems. The argument has been addressed by several scientific and regulatory bodies and in the last decade the debate in the field of endocrine disruptors has gained broad public attention. Much of the debate has been centred on the identification of endocrine disrupting chemicals (EDC), mechanisms of action, testing methods and non-monotonic dose-response relationships and low dose effects. In May 2016, a consensus statement regarding the identification of chemicals acting as endocrine disruptors, was reached in an open discussion by the scientific community from EU, USA and Japan (Solecki et al., 2016). The scientific community has agreed that the identification of EDC is a hazard identification procedure based on the WoE evaluation of an adverse health effect and on an endocrine mode of action. It was also agreed that potency is not a relevant issue and should not be taken into account for identification. It is anticipated that the statement will have an important impact because, according to EU Regulations, substances with endocrine capacity have to be replaced under pesticide and biocide regulations.

6.1.2 Application of the TTC to Developmental and Reproductive Toxicants – Overview

The application of the TTC approach to reproductive toxicants has been questioned by several authors. Kroes et al. (2000) gathered substances related to reproductive toxicity from the Munro et al. (1996) database, followed by the selection of NOEL values by focusing strictly on developmental effects, as these were assumed to occur at the lowest level of exposure. A total of 81 NOEL values for developmental toxicity were selected and it was concluded that cumulative distribution does not significantly differ from Cramer Class III NOEL distribution, therefore stricter thresholds than those reported by Munro et al (1996) would not be justified for developmental toxicants. In the study, the endocrine disuprtion toxicity endpoint, related to oestrogenic chemicals arising from foods, was considered separately. It was concluded that based on the knowledge available at that time, exposure to such chemicals is expected to be very low and that the 1.5 μg/person/day threshold is conservative enough to protect from any adverse endocrine effect associated with chemicals emerging from food (Kroes et al., 2000).

Following the Registration, Evaluation, Authorisation and restriction of Chemicals (REACH) regulation (ECHA, 2012), where the TTC approach is considered as a tool for chemical safety assessment at tonnage levels triggering limited information on repeated dose toxicity and/or reproduction, Bernauer et al (Bernauer et al., 2008) put together a dataset of 91 industrial chemicals tested for fertility and developmental toxicity. NOEL values were extracted from EU Risk Assessment Reports (RAR). A threshold value of 90 µg/day and 60 µg/day was reported for fertility and developmental toxicity respectively. These values again confirmed the validity of Munro derived thresholds for the structural Class III chemicals. van Ravenzwaay et al (van Ravenzwaay, Dammann, Buesen, Flick, & Schneider, 2011; van Ravenzwaay et al., 2012) derived a threshold value for developmental and maternal toxicity in rats and rabbits of 480 and 240 µg/day respectively. The dataset originated from BASF internal database on oral developmental toxicity testing and from published literature, mainly for pesticides and biocides, but also for veterinary products. A drawback of the above analyses was the small size of the datasets. Therefore, the work done by Laufersweiler at all (Laufersweiler et al., 2012) aimed to expand the dataset to give more robustness to thresholds related to Cramer structural classes. The dataset was compiled from previous studies and results from reproductive and developmental toxicity studies were included. Derived threshold values for each class, namely Class I, II and III, confirmed the conservativeness of the Munro et al (1996) thresholds in relation to the reproductive and developmental toxicity.

In a recent analysis by Stanard et al. (Stanard et al., 2015), a dataset of approved anticancer molecules was analysed to derive health-based thresholds for developmental and reproductive toxicity. A threshold value of 6 µg/day for anticancer compounds was suggested in cases where there is no assumption of hormone modulation activity. In cases where effects on hormones are expected or mode of action is unknown, it was suggested that different thresholds should be applied. That is 6 µg/day, 1 µg/day and 3 µg/day for reproduction/fertility, for developmental toxicity and for a combined developmental/reproductive toxicity respectively. It should be emphasised that the analysis presented included anticancer molecules with hormone (oestrogen) modulation activity, designed to be of high potency in order to have expected healthcare

outcomes. The potencies of such chemicals are presumably higher than those of environmental xenobiotics and closer to potencies of endogenous hormones. It is emphasised that the TTC approach should not be applied to substances with steroid structures.

6.1.3 EFSA and the Threshold of Toxicological Concern

EFSA applies the TTC approach for the risk assessment of flavouring substances and metabolites arising from pesticide applications. Improved technology allows for detection of more chemical substances present at low levels in food and feed, and some of them require risk assessment. Chemical substances present in food and feed include naturally occurring and synthetic substances and their relative breakdown products. These compounds usually lack toxicity data.

The application of the TTC approach has been so far applied to the risk assessment of chemicals present at low levels, where negligible exposure is expected. As contaminants are considered to be present at low levels, the application of TTC offers a way forward for the risk assessment in cases where there is lack of toxicological data. Accordingly, the EFSA's Scientific Committee was assigned to question the relevance and reliability of the TTC approach as a tool for providing scientific advice about possible human health risks from low-level exposures (EFSA, 2012b). Under the scope of the evaluation of the reliability and conservativeness of the TTC approach, it was also requested to review the applicability of the TTC for the application to developmental and reproductive toxicants.

6.1.4 Classification of Chemicals for Reproductive Toxicity

The Regulation (EC) No 1272/2008 on *Classification, labelling and packaging of substances and mixtures* (CLP), sets in place criteria to classify chemicals and their mixtures according to the identified hazards. This further allows appropriate hazard labelling upon which safety measure can be identified. The Regulation entered into force on 1 June, 2015 and it amends the *Dangerous Substance Directive 67/548/EEC14 (DSD),* the *Dangerous Preparations Directive 1999/45/EC15 (DPD)* and *Regulation (EC) No 1907/2006 (REACH)*. However, when data for

this analysis were provided, the old *Directive 67/548/EEC14 (DSD)* was still in place and the then valid classification system was applied and is explained below.

In the case of reproductive toxicity, the classification system identifies two types of adverse effects as defined in Guidance on the Application of the CLP Criteria (ECHA, 2015):

(a) Adverse effects on sexual function and fertility

Any effect of substances that has the potential to interfere with sexual function and fertility. This includes, but is not limited to, alterations to the female and male reproductive system, adverse effects on onset of puberty, gamete production and transport, reproductive cycle normality, sexual behaviour, fertility, parturition, pregnancy outcomes, premature reproductive senescence, or modifications in other functions that are dependent on the integrity of the reproductive system.

(b) Adverse effects on development of the offspring

Developmental toxicity includes, in its widest sense, any effect which interferes with normal development of the conceptus, either before or after birth, and resulting from exposure of either parent prior to conception, or exposure of the developing offspring during prenatal development, or postnatally, to the time of sexual maturation. However, it is considered that classification under the heading of developmental toxicity is primarily intended to provide a hazard warning for pregnant women, and for men and women of reproductive capacity.

Therefore, for pragmatic purposes of classification, developmental toxicity essentially means adverse effects induced during pregnancy, or because of parental exposure. These effects can be manifested at any point in the life span of the organism. The major manifestations of developmental toxicity include (1) death of the developing organism, (2) structural abnormality, (3) altered growth, and (4) functional deficiency.

Three categories are designated for reproductive toxicity effects, with Category 1 indicating that the chemical is *known* to cause adverse effects, Category 2 includes substances that *should be*

regarded as toxicants and for Category 3, substances that cause *concern* are considered. Categories are defined according to the level of evidence of the adverse effects. Inside each category, adverse effects, as explained above are identified, and the following risk phrases are suggested: for Category 1 and 2 R60 (May impair fertility) and R61 (May cause harm to the unborn child) and for Category 3 R62 (Possible risk of impaired fertility) and R63 (Possible risk of harm to the unborn child) apply. Based on the presented classification substances were selected for the analysis.

6.2 Aim

The aim of the investigation reported in this chapter was to classify substances for potential hazard concern, according to the Cramer classification and further analyse the distribution of NOEL values to derive the 5th percentile NOEL specific for the reproductive and developmental toxicants. This further allowed the evaluation of the conservativeness of structural thresholds proposed by Munro and colleagues (1996) for this specific and sensitive endpoint.

Results from this work have been used by the EFSA TTC working group for the development of a scientific opinion about the possible uses of the TTC approach to substance under EFSA remit.

6.3 Data and Methods

A list of 99 classified developmental toxicants and 65 classified reproductive toxicants was developed by the RIVM (the Netherlands) and made available (on a restricted basis) via the EFSA Working Group on TTC, in September 2010. The list included substances that are in the European Union classified for reproductive and/or developmental toxicity. Substances classified according to Directive 67/548/EEC for developmental toxicity (category 1, 2 or 3) or effects on sexual function and fertility (category 1, 2 or 3) were selected. The compounds' common name, CAS numbers and data on oral NOEL were available for some compounds.

In order to perform the analysis on the reliability of the current structural thresholds the author of the thesis performed the following steps: Chemical structures based on structure name and

CAS as provided in the files mentioned above were generated first and then stored in a file format that could be processed through Toxtree.Cramer classification needed to be applied. Toxtree (v 2.1.0), a computational tool for Cramer classification, was utilised for grouping of chemicals into hazard classes. Finally, distribution analysis of developmental and reproductive toxicological data was performed to derive the 5th percentile NOEL.

6.3.1 Construction of the Dataset

From the reproductive dataset, 6 inorganic compounds were removed and 33 unidentified (confidential) compounds were neglected. From the developmental dataset, the following were removed: 11 inorganics, 3 confidential compounds. In both datasets, salts were cleaned by removing the cationic part and the charge on the anion. Chemical structures for the remaining compounds, from the two sets, were not available. These were searched based on the CAS numbers and chemical common names using freely available web tools ChemSpider, available at http://www.chemspider.com/ and PubChem available at https://pubchem.ncbi.nlm.nih.gov/. SMILES were downloaded and stored in an Excel file based on which by utilising MarvinSketch structures were drawn and stored in a specific structural file, a .smi file.

6.3.2 Cramer Classification

The analysis was performed on 85 developmental toxicants (chemicals classified with risk phrases R61 or R63) and 54 fertility toxicants (chemicals classified with EU risk phrases R60 or R62).

The Cramer classification scheme, implemented in the Toxtree software (v 2.1.0),[6] was applied to the two set of compounds, namely the developmental and reproductive toxicants.

[6] http://toxtree.sourceforge.net/

6.3.3 Cumulative Distribution Analysis

Toxicological data were extracted from the original tables provided to the EFSA TTC working group by the RIVM (Potency database for substances with adverse effects on sexual function and fertility and Potency database for developmental toxicants) (Muller et al., 2012). However, NOEL overall data were not available for all substances. In the list of chemicals provided, the toxicological data include the type of study, species, route of administration, dose calculated to cause an increase incidence of 10% of a response ($ED_{10)}$, effect for Lowest Observed Adverse Effect level (LOAEL), and NOEL overall.

Considering only the NOEL overall data, 61 NOEL values were available for the developmental dataset and 41 for the fertility dataset. These were merged and then the cumulative distribution of NOEL values per Cramer Class was plotted using Microsoft excel. Then 5^{th} percentile NOEL values were derived based on the empirical distribution, following by the calculation of the human exposure threshold values for EFSA.

The results were further compared with TTC values proposed by Munro et al. 1996 for each Cramer Class, to check whether the proposed TTC levels would be protective for EFSA toxicants.

Microsoft Excel was used to calculate and plot the cumulative distribution curves presented in Figure 6.3 and to calculate the percentiles. Data for NOEL reproductive and developmental toxicants were combined.

The distribution curve for Class II was disregarded, as only four toxicological values were available.

6.4 Results

6.4.1 Cramer Classification

The aim of this chapter was to assign the Cramer hazard class to compounds classified for developmental and reproductive toxicity. Cramer classes were assigned using Toxtree. The

distribution of chemicals between Cramer classes is shown in the Figure 6.1 and Figure 6.2 below for the developmental and reproductive toxicants respectively. Most chemicals were classified for Cramer hazard Class III - 83% and 80% for developmental and reproductive toxicity respectively. 15% from both datasets were attributed to Cramer hazard Class I and only 4 compounds from both datasets were Cramer hazard Class II. The distribution patterns, among the three classes resembles very well the distribution of compounds from the reference Munro dataset having 73% of chemicals in Class III and 22% of chemicals in Class I.

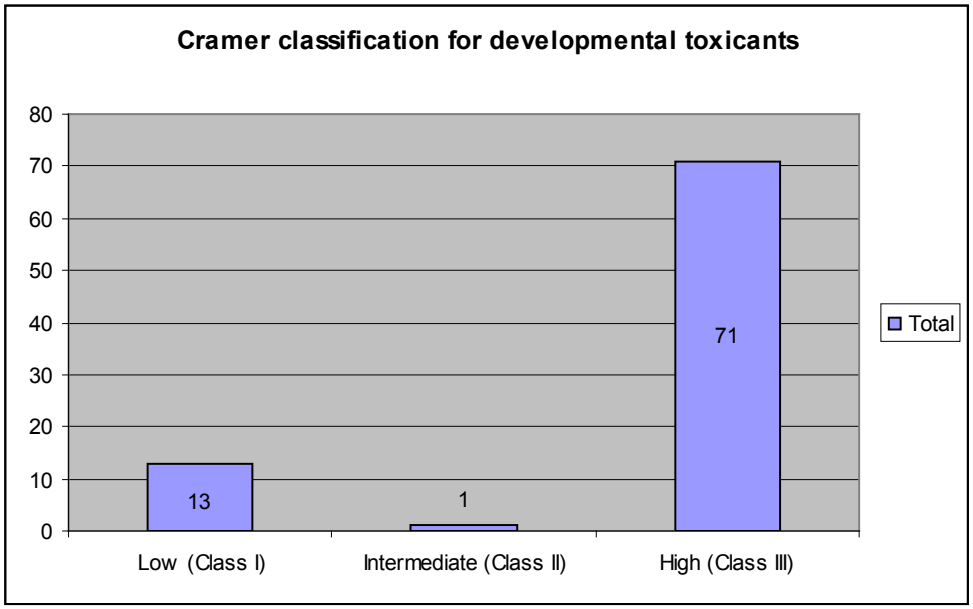


Figure 6.1. Distribution of 85 developmental toxicants among Cramer classes

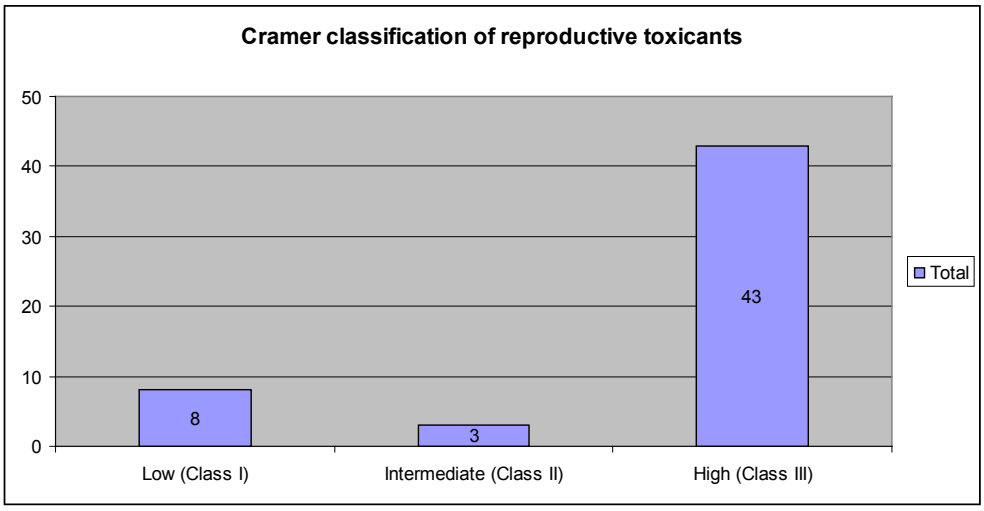


Figure 6.2. Distribution of 54 reproductive toxicants among Cramer classes

6.4.2 *Distribution Analysis*

For the calculation of the 5th percentile NOEL, chemicals from both datasets were combined, as only 102 NOEL values were available in the datasets. For the derivation of the 5th percentile NOEL organometal compounds were retained. Then, in the next step the empirical cumulative distribution of NOEL values per Cramer class was plotted. Subsequently theoretical log normal cumulative curves were fitted to the empirical data although further calculations were based only on empirical data. The distribution of chemicals belonging to Class II was not analysed, as there were too few chemicals (4). As shown in Figure 6.3, the cumulative NOELs are clustered around the solid lines, which shows a clear separation between the Class I and Class III toxicants in terms of their NOEL values. Only in the case of low NOEL values an apparent overlap is observed in the Cramer Class I and III curves. The low values of NOEL overall were related to organometal compounds.

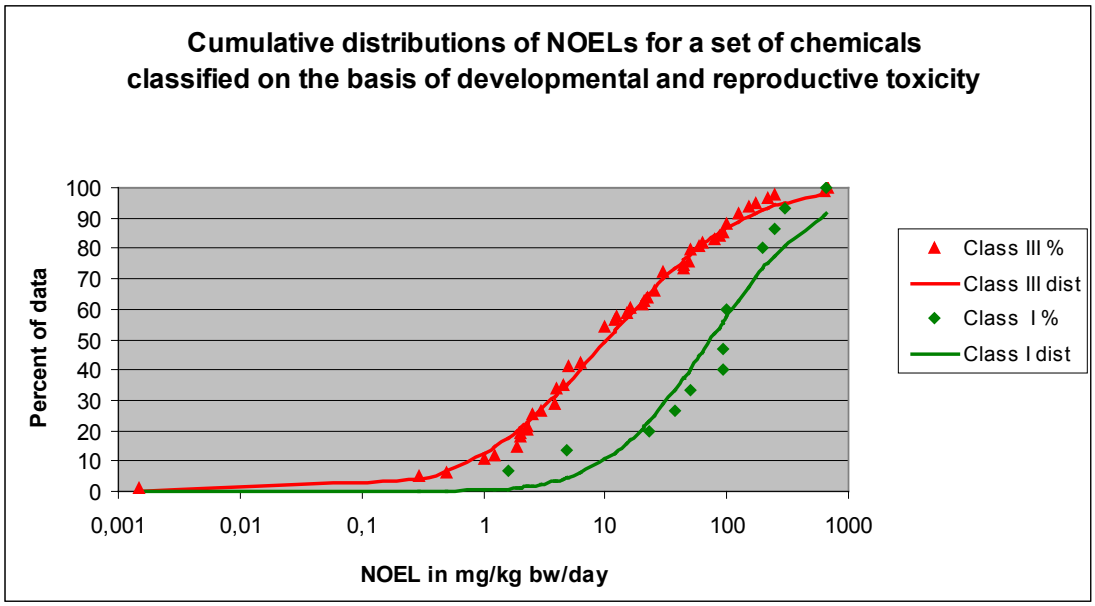


Figure 6.3. Empirical cumulative distribution of NOELs and log normally fitted cumulative distribution (solid lines)

The 5th percentile NOEL for each hazard Class was derived from the empirical cumulative distribution function (Table 6.1). These were then used to calculate the human exposure threshold, by multiplying the 5th percentile NOEL by 60 and dividing it by a safety factor of 100 as in the original Munro (1996) work.

Table 6.1. 5th Percentiles calculated for developmental and reproductive toxicants

Structural Hazard Class	No. of chemicals (developmental + reproductive NOEL)	5th percentile NOELS (µg/kg/day)	Human exposure threshold in (µg/day)
Class I	15	3840	2304
Class II	4		
Class III	83	550	330

Table 6.2. 5th percentiles and structural TTC as proposed by *Munro et al., 1996*

Structural Class	No. of chemicals	5th percentile NOELS (µg/kg/day)	Human exposure threshold in (µg/day)	Proposed TTC values (µg/kg/day)
Class I	137 (22%)	2 993	1 796	1 800
Class II	28 (5%)	906	544	540
Class III	446 (73%)	147	88	90

The calculated human exposure thresholds were compared with Munro's structural threshold (Table 6.2) to evaluate if the structural threshold would be protective for the EFSA toxicants dataset. The calculated human thresholds for Class III and Class I are higher than the Munro TTC values for the specific classes.

It can be concluded that the proposed structural TTC values for the Cramer classes I and III are protective for reproductive and developmental EFSA toxicants.

6.5 Discussion

Reproductive and developmental toxicity are very sensitive endpoints of high regulatory concern as they include adverse effects such as the inability to reproduce as well as malformations. The testing for these critical endpoints is requested within several regulatory frameworks, such as for the registration of plant protection products, under the REACH and biocides regulation and follows recommended testing procedures. Following the use of such products, toxic residues, for which no toxicity data are requested, may arise from metabolism of active ingredients or as impurities, and risk assessment has to be performed. A question arises

how to perform risk assessment when there is lack of toxicity data. Alternative approaches have been proposed and the TTC is one of them.

The aim of the present study was to assess the applicability of the TTC approach to developmental and reproductive toxicants relevant for EFSA, by applying Cramer classification and calculating the 5th percentile NOEL. A dataset, which included chemical substances that are in the European Union classified for reproductive and/or developmental toxicity (Muller et al., 2012), was made available for the analysis and results were used by EFSA working group on TTC to develop an opinion on the applicability of TTC to substances under EFSA remit also for the endpoint of reproductive and developmental toxicity (EFSA, 2012b).

The analysis confirmed that the distribution pattern of reproductive and developmental chemicals between the Cramer hazard classes resembles the distribution pattern of the compounds in the original Munro dataset very well, indicating that most compounds are classified as Cramer Class III. Further, by analysing the distribution of NOEL values and calculating the thresholds by deriving the 5th percentile NOEL from the empirical distribution and applying the safety factor of 100, it was confirmed, that the Munro thresholds currently in use, are protective enough also for EFSA's reproductive and developmental toxicants. This analysis has further confirmed the conclusions by several other authors, namely Bernauer (2008), Van Ravenzwaay (2011, 2012) and Laufersweiler (2011). All confirmed the validity of Cramer hazard classes specific threshold exposure values for reproductive and developmental toxicity, by analysing a broad spectrum of chemical classes including, industrial chemicals, pesticides and biocides.

EFSA (2012) in its opinion on the application of TTC to reproductive and developmental toxicants concluded that the TTC approach and current Cramer related thresholds remain valid for the reproductive endpoint. At this stage, a separate analysis including those chemicals acting via a hormone modulation activity was not possible, as mode of action of chemicals was not provided.

Reproductive and developmental toxicity should not be discussed in the absence of endocrine disruption. As demonstrated by Stanard (2015), hormone modulation activity occurs at lower levels, and lower threshold values were identified for such compounds. The threshold of 3 μg/day has been suggested by Stanard. That ED effects do occur at lower levels was also confirmed in epidemiological studies, which related environmental exposure to EDC with certain human diseases and disabilities (Vandenberg et al., 2012). Adverse effects at low doses itself however, do not implicate the non-existence of a threshold. At present, the existence of the effect threshold remains uncertain and leads to opposing opinions in the scientific community. The issue of non-monotonic dose response, low dose effects, remains a key research and regulatory challenge. It is recognised that current testing approaches, meaning testing at high doses, do not allow the acceptance or rejection of the existence of a threshold for ED chemicals. Therefore, only fundamental changes in chemical testing and safety determination would allow efficient protection of human health (Vandenberg et al., 2012).

As a result, the situation, may be similar to that for carcinogenicity, where carcinogens acting through a genotoxic mode of action have a lower threshold, the same could apply to reproductive and developmental toxicants where those chemicals acting via a hormone modulation mode of action would have lower or no threshold. It is also important to note that substances with endocrine disrupting capacity are conditionally exempted from exposure criteria, meaning that higher tier assays for these compounds are required also at lower production volumes (Tollefsen et al., 2014). For the compounds analysed in this study a mode of action was not provided, therefore possible oestrogenic effect could not have been considered. In the EFSA opinion (2012) it was concluded the TTC approach remains valid also for the assessment of untested chemicals, other than steroids, for what relates to the ED activity.

While the scientific community has agreed that potency is not a relevant issue for the identification of ED it was also anticipated that the non monotonic dose response and low dose issue will not be resolved in the near future. However, once the agreement regarding the low

dose effect due to non-monotonic dose response is reached possible effects on the application of TTC for this specific endpoint will need to be revised.

To contribute to the identification of chemicals with ED potential, Adverse Outcome Pathways (AOP) are being developed under the IATA strategy. An AOP represents existing knowledge concerning a series of events linking an initiating event at the molecular level (MIE) and the cascade of intermediate or key events (KEs) at the subcellular, cellular, tissue and organ level, that lead to a specific adverse outcome (AO) at the individual or population level. In this sense it provides a biological context, to support non testing methods under WoE approach (OECD, 2013; Tollefsen et al., 2014).

7 Discussion

In this final chapter, a short overview providing a summary and conclusions of the research is presented. Conclusions are discussed in the context of future work, in particular related to concerns raised throughout the work. It is noted that the TTC approach is being proposed for applications under several regulatory frameworks for different groups of chemicals, even for chemical mixtures and in the area of ecotoxicology. However, TTC should not be seen as a measure to omit data where specific requirements exist. It is the author's opinion, that future work should be centred on the limitations of the uses of the approach, such as exclusion groups and the verification of the underlying methods and data. In parallel to the ongoing research, the development of guidance and tools to support the end user should be provided.

7.1 Summary of the Findings of this Thesis

Due to an ever-increasing need for regulatory chemical safety assessment and the ability of more sophisticated analytical methods to detect xenobiotics to which individuals are exposed at low levels, alternative, specifically non testing - *in silico*, methods are considered in the context of WoE approach for chemical risk assessment. The TTC approach has been recognised as a valid approach in cases where low exposure levels are expected and where toxicological data are limited or even not available. TTC could also support the waiving of toxicological data where applicable within the regulatory context, such as REACH and BPR. The work presented in this thesis has focused on the applicability of *in silico* methods to support the application of TTC approach to different chemical classes and toxicity endpoints through the identification of genotoxic substances and development of improved databases.

The first chapter introduced the risk assessment process and explains where under the risk assessment paradigm the TTC approach could be applied. It further describes TTC and elaborates how the threshold values for different hazards were derived. It defines key issues: the uncertainty factors, the point of departure and the calculation of margin of safety, latter being an alternative approach to TTC for the assessment of carcinogens. The Cramer classification, a key concept in the TTC approach is presented. The TTC approach is given a context by an

overview of its historical development, thus allowing for a better understanding of the reasons and consequences of regulatory decisions in the past. It also explains where in the regulatory framework the incentives for the development of *in silico* tools lie. The first chapter concludes with a description of prospective usage of *in silico* tools and methods in the application of TTC.

Genotoxicity is one of the first toxicities to be assessed in any chemical safety assessment, including TTC. A threshold value, a virtually safe dose, is established for exposures to chemicals which are not part of Cohort of Concern (COC) but contain a structural alert for genotoxicity. ***Therefore, the second chapter introduced the endpoint of genotoxicity and evaluates non-testing methods to predict it.*** It is further described how a set of nitrobenzenes with Ames mutagenicity data was extracted and compiled using a software tool that integrates a searchable database with results from toxicity studies and predictive models. Extracted data included the nitrobenzene structure, identifiers such as SMILES, CAS and chemical common name, and Ames test results from all *Salmonella* strains tested. A table comprising 252 nitrobenzenes with Ames test results was created and data were extracted into structural files. In the second part of Chapter 2, predictions of genotoxicity were obtained from seven commercial and freely available software tools, used by industry and regulatory authorities.

The predictions obtained were evaluated for predictive ability according to Cooper statistics. The analysis demonstrated a good ability of expert knowledge systems to identify a structural alert for genotoxicity. However, these were less successful in correctly identifying the non-genotoxic nitrobenzenes. In this respect CAESAR, which combines statistical approach and a mutagenicity structural alert profiler proved to be the best performing tool. Chemoinformatic tools were applied to identify possible physicochemical parameters contributing to the genotoxic activity of nitrobenzenes. Some compounds negative in Ames test seem to be correlated with high lipophilicity values. Data collected for the analysis presented in Chapter 2 were used by (Kharchevnikova, Blinova, Dobrynin, & Jurkov, 2012) from the Russian National Academy of Science to derive a model for the prediction of nitrobenzenes. There, it was demonstrated that some negative results could be explained by the stability of metabolites. The

research presented in the two chapters demonstrated the validity of the QSAR methods in the sense of the ability to identify a structural alert for genotoxicity thus generating the information. An appreciated asset of expert knowledge tools was demonstrated to be the application of mechanistic information regarding genotoxicity. Although the overall result was that expert knowledge tools are over predictive, the conservativeness built in is fit for regulatory purposes. As for statistical tools the reliability of predictions should always be assessed in the context of the applicability domain of the QSAR model.

Chapter 3 further elaborated the results of Chapter 2 in order to expand the research and develop a case study. The case study shows the potential applicability of the non-testing methods for the risk assessment of pesticides, the risk assessment of which is under EFSA remit. The plant protection product selected for the study, also a nitrobenzene, is a plant growth regulator, sodium 5-nitroguaiacolate. It is included in the list of approved active substances, under regulation EC No 1107/2009. The first part of the chapter introduces mutagenicity test results of the active substance. Then the same approach as in Chapter 2 is applied to predict genotoxicity (identify a SA for genotoxicity). Based on the mechanistic knowledge acquired from knowledge-base systems and based on chemical structural properties, in the second part of the chapter, chemical similarity, category formation and read across were utilised to predict the genotoxicity of sodium 5-nitroguaiacolate, a nitrobenzene. The approach proved the robustness of read across as predictions for sodium 5-nitroguaiacolate were in line with the Ames test results, that is they are negative.

The first part of the work analysed the possibilities to utilise *in silico* methods to support the application of the first step in the TTC decision tree, that is to identify SA for genotoxicity. The following three chapters, namely the 4th, 5th and 6th, tried to demonstrate the applicability of Cramer hazard classes related thresholds to cosmetic ingredients, biocide active substances, and developmental and reproductive substances.

Chapter 4 focused on the applicability of the TTC approach to cosmetic ingredients. The research was based on the work performed in the initial phase of the COSMOS Project, the

results of which have been used by the EU Scientific committees, the SCCS, SCHER and SCENIHR (2012) to produce an opinion on the application of TTC with the focus on cosmetics. The work presented applies non-testing methods to estimate the representativeness of the Munro database for cosmetic ingredients. Further, Cramer classification and the NOEL distribution analysis were performed for the non-cancer COSMOS TTC dataset. COSMOS was, at that time, an on-going project and since then the toxicological dataset has undergone an extensive review and new results are available since the end of the project in December 2015. The final results confirmed the validity of Cramer Class I and III thresholds, where for Class II an intriguing lower threshold value was calculated. Also classes of chemicals under-represented in the Munro dataset were identified and further added to the Munro dataset. For this extended dataset new Munro thresholds were derived, which are still in the same range of previously published thresholds, but less conservative (Yang et al., 2016).

Similarly, in Chapter 5 the applicability of TTC approach to biocides active substances was analysed. Biocide active substances were available for research as part of an internal collaboration within the Institute for Health and Consumer Protection (IHCP) at the JRC. In the first part of the chapter, data were curated and structure files were prepared. Chemoinformatics tools were applied to demonstrate the representativeness of the Munro database in terms of physicochemical parameters and structural representations. It was determined that to some extent, the set of biocides differs from the original Munro dataset in terms of physicochemical properties, indicating biocides might be more bioavailable than the class of chemicals covered by the Munro dataset. Missing groups of chemicals were identified, among them peroxides, surfactants, and thiazole compounds. Subsequently, a smaller set of compounds (77) was identified, with toxicological data collected from public sources. Based on this collection, Cramer classification and distribution analysis were applied to perform a preliminary study aimed at demonstrating the validity of currently applied threshold values for biocides active substances. The distribution of chemicals, of the whole biocides dataset and the TTC biocides dataset, between the Cramer classes is similar to the distribution of the Munro dataset. That is, the majority of compounds are placed in Class III and only 14 chemicals are classified as Cramer

Class I. The calculated threshold for Class I, which was below the Munro's threshold, was not considered reliable. Instead for Class III the threshold is slightly below the Munro threshold, but still in its range. At this time, based on the presented analysis, firm conclusion regarding the protectiveness of the threshold of Class I compounds cannot be drawn especially due the limited NOEL dataset. However the analysis indicates that especially for Class III the threshold is valid and that TTC is a conservative and applicable approach also for biocide active substances.

In Chapter 6, a list of chemical compounds with related toxicological data was utilised to verify the applicability of the currently available threshold for chemicals classified for developmental and reproductive toxicity under the Dangerous Substances Directive (67/548/EEC) and relevant for EFSA. The list of compounds was defined in a working document provided by RIVM to the EFSA TTC Working Group. Chemoinformatics tools were applied to construct structure files that allowed for the application of Cramer classification with Toxtree. Toxicological data provided were curated to perform the NOEL distribution analysis. It was demonstrated that the distribution of chemicals between the three Cramer classes resembles the classification of compounds from the Munro dataset closely, meaning that most compounds are classified in Class III. Only four compounds were identified as Cramer class II and these were excluded from further analysis. The distribution analysis and the derivation of the 5th percentile NOEL for compounds proposed TTC values, for the Cramer classes I and III, for reproductive and developmental EFSA toxicants. These were above the currently applicable thresholds. These results were used by EFSA to develop an opinion on the application of TTC (EFSA (2012).

7.2 Conclusions

The Threshold of Toxicological Concern approach is suggested for application to several groups of chemicals as well as to specific endpoints. As the approach is suggested to be applicable where no substance specific toxicological data are available, the combination of non-testing methods and tools for prediction of genotoxicity developed in the Chapter 2 and 3 is understood in the context of the WoE approach and is suggested for the identification of genotoxic

nitrobenzenes. The expert knowledge tools were able to identify structural alerts and provide mechanistic interpretation of genotoxicity. This mechanistic knowledge can be used further to construct a category of structurally related compounds able to induce the same mechanism of genotoxicity. The category can be used for read-across to fill a data gap in genotoxicity for a non-tested compound. An approach that combines several non-testing approaches, such as tools to identify structural alerts, calculate basic physico-chemical properties, collect mechanistic knowledge about the structural alert and similar compounds, as well the knowledge on metabolites is suggested. The results from such evaluation would provide the expert with relevant knowledge to decide upon the genotoxicity of a compound and apply the relevant exposure threshold (TTC).

The assumption that the thresholds derived from Munro dataset which was based on non-cancer repeat dose toxicity studies and calculated for structural Cramer classes, and is currently applicable, was tested with regard to cosmetic ingredients and biocides, and to developmental and reproductive toxicants in Chapters 4, 5 and 6. In all three chapters, Cramer classification was assigned with Toxtree and toxicological data were used to derive the 5th percentile NOEL related to the Cramer hazard classes. For chemicals from Chapter 4 and 6, the applicability of Cramer Class III threshold was supported by the fact that the threshold values as currently applied are below the calculated threshold for the two groups of chemicals. In the case of biocides, Chapter 5, a slightly lower threshold was derived. This could be explained by the fact that biocides active compounds were analysed and that these chemicals are designed to be active. Further, no quality check of the data was possible at this stage and the NOEL selected do not necessarily represent the lowest NOAEL. Additionally, carcinogenic chemicals were retained in the dataset to keep the threshold conservative as the NOEL dataset was very limited. For Class I chemicals, different results were observed. The Class I threshold is not protective for cosmetics and biocides but is protective for developmental and reproductive toxicants.

The non-protectiveness of the thresholds could be explained with misclassifications, that is, by the inability of the Cramer classification to efficiently discriminate between Class I and III

chemicals, taking into account that a chemical with a lower NOEL is more toxic. It could be also explained by the sole fact that the TTC is not protective for Class I chemicals. Additionally, in the case of biocide active substances, the number of compounds classified in Class I is very limited (14 compounds). Therefore, it is difficult to draw firm conclusions on the reliability of the threshold for this set of chemicals. In all cases there were only a small number of Class II chemicals identified which did not further support the evaluation of the applicability of the threshold value to this group. As the classification of a chemical in a hazard class is one of the key steps in the application TTC, a potentially incorrect classification of a chemical is then reflected in the wrong threshold selected, thus leading to the wrong judgement about the hazard of a chemical and inappropriate risk conclusions.

The usefulness of chemoinformatics tools is demonstrated in Chapters 4 and 5 where these tools were broadly applied to compare the Munro dataset with cosmetic ingredients and biocide active substances in terms of the substructure representations and physicochemical parameters. It was demonstrated that certain groups of chemicals (surfactants, peroxides) are not represented or are under-represented in the Munro dataset. It was also shown in Chapters 4 and 5 that the biocide active substances and cosmetic ingredients dataset differ in some physiochemical properties in comparison to the Munro dataset.

7.3 Prospects for Future Work

7.3.1 Reflections on Cramer Classification

Cramer classification implemented in Toxtree has been applied throughout the thesis for hazard classification of different groups of chemicals. Since the Cramer classification was published (1976), advances in understanding the way chemicals exert toxicity provide an additional source of information to classify chemicals for their toxic hazard. This is an incentive for reviewing the Cramer classification and attempts to improve the rules to classify chemicals with unknown toxicological profiles. Several studies confirmed the validity of classification of Class III and further support their use (Lauferswiller 2011, Tluczkiewicz 2011), which is in line with the work

presented. However, for Class I chemicals Cramer classification seems to be over conservative in most cases and it does not seem to work for Class II compounds. Therefore, suggestions to update and review the classification scheme are reported, especially related to fragrances (Bhatia et al., 2015; Roberts, Api, et al., 2015; Schnabel & Taylor, 2015).

The improvement of Cramer scheme rules, based on a specific group of chemicals, might adversely affect the classification of chemicals from other groups. Therefore, a researcher should bear in mind that any proposed refinements should be extensively validated on databases from a broad range of chemical classes to understand the effects on currently applied thresholds. Since the definition of the Cramer scheme, new understandings have been gained in several areas of toxicology and the many new possibilities are offered by chemoinformatics tools to derive structure-specific data in the form of physico-chemical descriptors and of mechanistic structural alerts. Instead of trying to "fix" the Cramer scheme, a better approach may be to exploit that new scientific and technological development to define a new classification system. New computational approaches, including big data analysis, machine learning and data mining, could be used to identify specific patterns. For example, in physico-chemical descriptors related to toxicity and derive decision algorithms.

Another alternative would be to move away from grouping of chemicals into hazard classes and consider the development of computational tools and methods for the identification of a molecular initiating event. A MIE is defined by Allen et al (Allen, Goodman, Gutsell, & Russell, 2014) as the initial interaction between a molecule and a biomolecule or biosystem that can be causally linked to an outcome via a pathway. The evaluation presented in the thesis, demonstrated that computational tools are applicable for the identification of structural alerts connected with a molecular initiating event related to genotoxicity. In this same way SA for other MIEs connected with other adverse outcomes could be developed.

From the regulatory point of view, it is considered that effort should be also placed in the identification of chemicals or chemical groups to which the TTC approach is not applicable and

in the development of tools and protocols for the identification of chemicals to which the TTC can be applied.

7.3.2 Reflections on Thresholds for Biocide Active Substances

The present study with biocide active substances is a preliminary study aimed to assess the validity of currently applied threshold values for those substances. It did not attempt to set new threshold values for biocides; instead, it wanted to explore the possible application of the TTC approach to biocides. In the study, some drawbacks were identified. The study identified that, to some extent, the set of biocides differs from the original Munro dataset, indicating biocides might be more bioavailable than the class of chemicals covered by the Munro dataset. In addition, missing groups of chemicals were identified - among them peroxides, surfactants, and thiazole compounds.

Toxicity data for a subset of biocide active compounds were gathered from public sources, and for that reason they are very limited. The NOEL values gathered might not necessarily represent the regulatory established lowest NOAEL for a chemical, therefore giving additional conservativism to the results. In addition, no further quality evaluation of the data was performed at this stage. The study is considered to be conservative and no additional uncertainty factors were applied, for example due to the small dataset, when calculating the thresholds for human exposure. Even when considering all the drawbacks of the study, the final threshold value for Class III is in the range of the current threshold, which additionally supports its validity. As for Class I, due the limited size of the dataset the reliability of the threshold is questionable.

These findings give motivation for further research and show the necessity and relevance of open access data to build a bigger dataset in order to give more robustness to the analysis. This could be achieved, for example, by extracting data from registration dossiers. New studies demonstrating that the lowest NOEL selected is influenced by the study design and availability of studies (Zarn et al., 2015), should also be considered when extracting toxicological data for TTC calculation.

Biocides can be found among cosmetic ingredients as for example surfactants and preservatives. Therefore the results from the analysis of the applicability of TTC to cosmetic ingredients are to some extent applicable to the biocides analysis, for example to understand the effect of missing chemical groups on the threshold. Now that the final dataset for non-cancer COSMOS TTC is available (http://cosmosdb.cosmostox.eu/) the toxicological data could be extracted to improve the biocides TTC dataset.

7.3.3 Concluding Remarks

TTC is a tool mainly applied to support the risk assessment of chemicals to which individuals are exposed at low doses. It can be applied if two prerequisites are met, good exposure data and information about the chemical structure. The study that was undertaken for this thesis analysed several aspects of the TTC approach and leads to the following conclusions about the application of this approach.

By deriving threshold values, for new datasets of chemical compounds, of regulatory interest, it was confirmed that the TTC is a general pragmatic and conservative approach to risk assessment, for low exposures, in the absence of chemical specific toxicity data.

There is much interest in updating the dataset for TTC (that is to expand the approach to other sectors – cosmetics, biocides) and for the refinement of the Cramer classification tree. Therefore it is considered that the development of datasets should be transparently reported. Extensions of the original Munro dataset should be evaluated also in the sense of the effect on the current thresholds, as this should remain conservative, to maintain the general applicability of TTC as the approach in risk assessment of chemicals.

As TTC approach is a probability tool its strength lies in the fact that is conservative and is author's opinion that this strength should be reinforced.

Cramer classes could be revised, updated or replaced by new mechanistic understanding. However, it is considered, based on the presented evaluation that the current Cramer classification scheme is fit for regulatory purposes, especially for Cramer Class III chemicals.

Eventual updates of the Cramer classification would need to undergo extensive evaluations of the impact on current thresholds.

8 References

Adler, S., Basketter, D., Creton, S., Pelkonen, O., Van Benthem, J., Zuang, V., et al. (2011). Alternative (non-animal) methods for cosmetics testing: Current status and future prospects-2010. *Archives of Toxicology*, *85*(5), pp. 367–485.

Allen, T. E. H., Goodman, J. M., Gutsell, S., & Russell, P. J. (2014). Defining molecular initiating events in the adverse outcome pathway framework for risk assessment. *Chemical Research in Toxicology*, *27*(12), pp. 2100–2112.

Ashby, J., & Tennant, R. W. (1991). Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. *Mutation Research/Reviews in Genetic Toxicology*, *257*(3), pp. 229–306.

Ball, N., Cronin, M. T. D., Shen, J., Blackburn, K., Booth, E. D., Bouhifd, M., et al. (2016). Toward good read-across practice (GRAP) guidance. *ALTEX*, *33*(2), pp. 149–166.

Barlow, S. (2005). *Threshold of Toxicological Concern (TTC): A Tool for Assessing Substances of Unknown Toxicity Present at Low Levels in the Diet*. ILSI Europe.

Bassan, A., Fioravanzo, E., Pavan, M., & Stocchero, M. (2011). *Scientific report submitted to EFSA on Applicability of physicochemical data , QSARs and read-across in Threshold of Toxicological Concern assessment*. Vicenza: S-IN Soluzioni Informatiche.

Beland, F. A., & Poirier, M. C. (1989). DNA Adducts and Carcinogenesis, in: A. E. Sirica (Ed.), *The Pathobiology of Neoplasia*, pp. 57–80. Boston, MA: Springer US.

Belanger, S. E., Sanderson, H., Embry, M. R., Coady, K., Dezwart, D., Farr, B. A., et al. (2015). It is time to develop ecological thresholds of toxicological concern to assist environmental hazard assessment. *Environmental Toxicology and Chemistry*, *34*(12), pp. 2864–2869.

Benigni, R. (2014). Predicting the carcinogenicity of chemicals with alternative approaches: recent advances. *Expert Opinion on Drug Metabolism & Toxicology*, *10*(9), pp. 1199–

1208.

Benigni, R., & Bossa, C. (2006). Structural Alerts of Mutagens and Carcinogens. *Current Computer - Aided Drug Design*, *2*(2), pp. 169–176.

Benigni, R., & Bossa, C. (2011). Mechanisms of chemical carcinogenicity and mutagenicity: A review with implications for predictive toxicology. *Chemical Reviews*, *111*(4), pp. 2507–2536.

Benigni, R., Bossa, C., Jeliazkova, N., Netzeva, T., & Worth, A. (2008). *The Benigni / Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree*. Joint Research Centre.

Bernauer, U., Heinemeyer, G., Heinrich-Hirsch, B., Ulbrich, B., & Gundert-Remy, U. (2008). Exposure-triggered reproductive toxicity testing under the REACH legislation: A proposal to define significant/relevant exposure. *Toxicology Letters*, *176*(1), pp. 68–76.

Bhatia, S., Schultz, T., Roberts, D., Shen, J., Kromidas, L., & Marie Api, A. (2015). Comparison of Cramer classification between Toxtree, the OECD QSAR Toolbox and expert judgment. *Regulatory Toxicology and Pharmacology*, *71*(1), pp. 52–62.

Bitsch, A., Jacobi, S., Melber, C., Wahnschaffe, U., Simetska, N., & Mangelsdorf, I. (2006). REPDOSE: A database on repeated dose toxicity studies of commercial chemicals-A multifunctional tool. *Regulatory Toxicology and Pharmacology*, *46*(3), pp. 202–210.

Blackburn, K., Stickney, J. A., Carlson-Lynch, H. L., McGinnis, P. M., Chappell, L., & Felter, S. P. (2005). Application of the threshold of toxicological concern approach to ingredients in personal and household care products. *Regulatory Toxicology and Pharmacology*, *43*(3), pp. 249–259.

Błaziak, K., Danikiewicz, W., & Mąkosza, M. (2016). How Does Nucleophilic Aromatic Substitution Really Proceed in Nitroarenes? Computational Prediction and Experimental

Verification. *Journal of the American Chemical Society*, *138*(23), p. 7276−7281.

Boobis, A. (2015). Cancer thresholds, Cohort of Concern and other excluded substance groups. *Toxicology Letters*, *238*(2), p. S6.

Boobis, A., Budinsky, R., Collie, S., Crofton, K., Embry, M., Felter, S., et al. (2011). Critical analysis of literature on low-dose synergy for use in screening chemical mixtures for risk assessment. *Critical reviews in toxicology*, *41*(5), pp. 369–83.

Carthew, P., Clapp, C., & Gutsell, S. (2009). Exposure based waiving: The application of the toxicological threshold of concern (TTC) to inhalation exposure for aerosol ingredients in consumer products. *Food and Chemical Toxicology*, *47*(6), pp. 1287–1295.

Cassano, A., Raitano, G., Mombelli, E., Fernandez, A., Cester, J., Roncaglioni, A., et al. (2014). Evaluation of QSAR models for the prediction of ames genotoxicity: a retrospective exercise on the chemical substances registered under the EU REACH regulation. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev*, *32*(3), pp. 273–298.

Cheeseman, M. A. (2005). Thresholds as a unifying theme in regulatory toxicology. *Food additives and contaminants*, *22*(10), pp. 900–6.

Cheeseman, M. A., Machuga, E. J., & Bailey, A. B. (1999). A tiered approach to threshold of regulation. *Food and Chemical Toxicology*, *37*(4), pp. 387–412.

COC. (2014). *Defining a Point of Departure and Potency Estimates in Carcinogenic Dose Response*. Committee on Carcinogenicity.

Cooper, J. a, Saracci, R., & Cole, P. (1979). Describing the validity of carcinogen screening tests. *British journal of cancer*, *39*, pp. 87–89.

Cozigou, G., Crozier, J., Hendriksen, C., Manou, I., Ramirez-Hernandez, T., & Weissenhorn, R. (2015). The European Partnership for Alternative Approaches to Animal Testing (EPAA): promoting alternative methods in Europe and beyond. *Journal of the American*

Association for Laboratory Animal Science : JAALAS, *54*(2), pp. 209–13.

Cramer, G. M., & Ford, R. A. (1978). Estimation of toxic hazard - a decision tree approach. *Food and Chemical Toxicology*, *16*, pp. 255–276.

Cronin, M., & Madden, J. (2010). *In Silico Toxicology* (M. Cronin & J. Madden, Eds.). Cambridge: Royal Society of Chemistry.

Cronin, M. T. D., Jaworska, J. S., Walker, J. D., Comber, M. H. I., Watts, C. D., & Worth, A. P. (2003). Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environmental Health Perspectives*, *111*(10), pp. 1391–1401.

Damme, S. K., Ratte, H.-T., Hollert, H., Coors, A., Knacker, T., Rettinger, K., et al. (2011). ECOSM-a new joint project for assessing environmental risks of poorly soluble compounds used in cosmetics-project presentation. *Environmental Sciences Europe*, *23*(1), p. 30.

DG Environment, E. C. (2007). *Study on Impact of the implementation of Directive 98 / 8 / EC concerning the placing on the market of biocidal products*. DG Environment, European Commission.

EC. (2006). *Clarification on Substances Notified Under the BKC and DDAC Generic Headings in Annex OO of Commission Regulation EC No 2032/2003*. European Commision.

EC. (2008). *Draft Opinion on Use of the Threshold of Toxicological Concern (TTC) Approach for the Safety Assessment of Chemical Substances*. European Commision.

ECHA. (2010). *Guidance on information requirements and chemical safety assessment Chapter R . 5 : Adaptation of information requirements*. European Chemicals Agency.

ECHA. (2012). *Guidance on information requirements and chemical safety assessment Chapter R.7c: Endpoint specific guidance*. European Chemicals Agency.

ECHA. (2015). *Guidance on the Application of the CLP Criteria*. European Chemicals Agency.

EFSA. (2005). Opinion of the Scientific Committee on a request from EFSA related to A Harmonised Approach for Risk Assessment of Substances Which are both Genotoxic and Carcinogenic. *The EFSA Journal*, *282*, pp. 1–31.

EFSA. (2008). *CONCLUSION ON PESTICIDE PEER REVIEW Conclusion regarding the peer review of the pesticide risk assessment of the active substances sodium 5-nitroguaiacolate, sodium o-nitrophenolate and sodium p-nitrophenolate*. EFSA Scientific Report.

EFSA. (2012a). Scientific Opinion on Evaluation of the Toxicological Relevance of Pesticide Metabolites for Dietary Risk Assessment 1. *EFSA Journal*, *10*(7), p. 2799.

EFSA. (2012b). 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), pp. 1–103.

EFSA, & WHO. (2016a). Review of the Threshold of Toxicological Concern (TTC) approach and development of new TTC decision tree. *EFSA Supporting Publication 2016:EN-1006*, pp. 1–50.

EFSA, & WHO. (2016b). *Outcome of a public consultation on the conclusions and recommendations of the EFSA – WHO workshop on the Threshold of Toxicological Concern approach European Food Safety Authority and World Health Organization*. European Food Safety Authority.

Ellison, C. M., Madden, J. C., Judson, P., & Cronin, M. T. D. (2010). Using in silico tools in a weight of evidence approach to aid toxicological assessment. *Molecular Informatics*, *29*(1-2), pp. 97–110.

Ellison, C. M., Sherhod, R., Cronin, M. T. D., Enoch, S. J., Madden, J. C., & Judson, P. N. (2011). Assessment of methods to define the applicability domain of structural alert

models. *Journal of Chemical Information and Modeling*, *51*(5), pp. 975–985.

EMEA. (2006). *Guideline on the limits of genotoxic impurities*. European Medicines Agency.

Enoch, S. J., & Cronin, M. T. D. (2010). A review of the electrophilic reaction chemistry involved in covalent DNA binding. *Critical reviews in toxicology*, *40*(8), pp. 728–748.

Enoch, S. J., & Cronin, M. T. D. (2012). Development of new structural alerts suitable for chemical category formation for assigning covalent and non-covalent mechanisms relevant to DNA binding. *Mutation Research - Genetic Toxicology and Environmental Mutagenesis*, *743*(1-2), pp. 10–19.

Escher, S. E., Tluczkiewicz, I., Batke, M., Bitsch, A., Melber, C., Kroese, E. D., et al. (2010). Evaluation of inhalation TTC values with the database RepDose. *Regulatory Toxicology and Pharmacology*, *58*(2), pp. 259–274.

EU. (2009). *REGULATION (EC) No 1107/2009 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC*. European Parliament and the Council of the European Union.

EU. (2012). *REGULATION (EU) No 528/2012 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 May 2012 concerning the making available on the market and use of biocidal products ("BPR")*. European Parliament and the Council of the European Union.

Feigenbaum, A., Pinalli, R., Giannetto, M., & Barlow, S. (2015). Reliability of the TTC approach: Learning from inclusion of pesticide active substances in the supporting database. *Food and Chemical Toxicology*, *75*, pp. 24–38.

Felter, S., Lane, R. W., Latulippe, M. E., Craig Llewellyn, G., Olin, S. S., Scimeca, J. A., et al. (2009). Refining the threshold of toxicological concern (TTC) for risk prioritization of

trace chemicals in food. *Food and Chemical Toxicology*, *47*(9), pp. 2236–2245.

Fioravanzo, E., Bassan, a, Pavan, M., Mostrag-Szlichtyng, a, & Worth, a P. (2012). Role of in silico genotoxicity tools in the regulatory assessment of pharmaceutical impurities. *SAR and QSAR in environmental research*, (March), pp. 37–41.

Frawley, J. P. (1967). Scientific evidence and common sense as a basis for food-packaging regulations. *Food and Cosmetics Toxicology*, *5*(January), pp. 293–308.

Fuart Gatnik, M., & Worth, A. (2010). *Review of Software Tools for Toxicity Prediction*. European Commission, Joint Research Centre.

Gocht, T., & Schwarz, M. (2014). *Towards the Replacement of in vivo Repeated Dose Systemic Toxicity Testing*. Coach consortium.

Gross, M. Y., Daginnus, K., Deviller, G., de Wolf, W., Dungey, S., Galli, C., et al. (2009). Thresholds of Toxicological Concern for Endocrine Active Substances in the Aquatic Environment. *Integrated Environmental Assessment and Management*, *6*(1), pp. 2–11.

Gupta, R. L., Saini, B. H. K., & Juneja, T. R. (1997). Nitroreductase independent mutagenicity of 1-halogenated-2,4-dinitrobenzenes. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*, *381*(1), pp. 41–47.

Gutsell, S., Hodges, G., Marshall, S., & Roberts, J. (2015). Ecotoxicological thresholds-practical application to an industrial inventory. *Environmental Toxicology and Chemistry*, *34*(4), pp. 935–942.

Hahn, S., Schneider, K., Gartiser, S., Heger, W., & Mangelsdorf, I. (2010). Consumer exposure to biocides--identification of relevant sources and evaluation of possible health effects. *Environmental health : a global access science source*, *9*(1), p. 7.

Hayes, A. W. (Ed.). (2008). *Principles and Methods of Toxicology*. CRC Press.

Hollnagel, H. M., Arvidson, K., Boobis, A. R., Barlow, S., Cronin, M. T., Felter, S. P., et al. (2016). Final report on the Development of a Non-Cancer Threshold of Toxicological Concern (TTC) Database to Support Alternative Assessment Methods for Cosmetics-Related Chemicals, in: *Supplement to Toxicological Sciences - The Toxicologist*, p. 349.

Hsu, C.-H., Stedeford, T., Okochi-Takada, E., Ushijima, T., Noguchi, H., Muro-Cacho, C., et al. (2007). Framework Analysis for the Carcinogenic Mode of Action of Nitrobenzene. *Journal of Environmental Science and Health, Part C*, *25*(2), pp. 155–184.

Ju, K.-S., & Parales, R. E. (2010). Nitroaromatic compounds, from synthesis to biodegradation. *Microbiology and molecular biology reviews : MMBR*, *74*(2), pp. 250–72.

Kalgutkar, A. S., Gardner, I., Obach, R. S., Shaffer, C. L., Callegari, E., Henne, K. R., et al. (2005). A Comprehensive Listing of Bioactivation Pathways of Organic Functional Groups. *Current Drug Metabolism*, *6*, pp. 161–225.

Kazius, J., McGuire, R., & Bursi, R. (2005). Derivation and validation of toxicophores for mutagenicity prediction. *J. Med. Chem*, *48*, pp. 312–320.

Kerns, E. H., & Di, L. (2008). *Drug-like Properties: Concepts, Structure Design and Methods from ADME to Toxicity Optimization*. Elsevier.

Kharchevnikova, N. V., Blinova, V. G., Dobrynin, D. A., & Jurkov, V. S. (2012). Analysis of the mutagenicity of nitrobenzenes in the Ames test by means of the JSM (John Stuart Mill) method, in: *15th International Workshop on Quantitative Structure-Activity Relationships in Environmental and Health Sciences*.

Kirkland, D., Aardema, M., Henderson, L., & Müller, L. (2005). Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens I. Sensitivity, specificity and relative predictivity. *Mutation research*, *584*(1-2), pp. 1–256.

Knudsen, T. B., Martin, M. T., Kavlock, R. J., Judson, R. S., Dix, D. J., & Singh, A. V. (2009). Profiling the activity of environmental chemicals in prenatal developmental toxicity studies using the U.S. EPA's ToxRefDB. *Reproductive Toxicology*, *28*(2), pp. 209–219.

Koster, S., Boobis, A. R., Cubberley, R., Hollnagel, H. M., Richling, E., Wildemann, T., et al. (2011). Application of the TTC concept to unknown substances found in analysis of foods. *Food and Chemical Toxicology*, *49*(8), pp. 1643–1660.

Koster, S., Leeman, W., Verheij, E., Dutman, E., van Stee, L., Nielsen, L. M., et al. (2015). A novel safety assessment strategy applied to non-selective extracts. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*, *80*, pp. 163–81.

Kroes, R., Galli, C., Munro, I., Schilter, B., Tran, L. A., Walker, R., et al. (2000). Threshold of toxicological concern for chemical substances present in the diet: A practical tool for assessing the need for toxicity testing. *Food and Chemical Toxicology*, *38*(2-3), pp. 255–312.

Kroes, R., Kleiner, J., & Renwick, A. (2005). The threshold of toxicological concern concept in risk assessment. *Toxicological Sciences*, *86*(2), pp. 226–230.

Kroes, R., Renwick, A. G., Cheeseman, M., Kleiner, J., Mangelsdorf, I., Piersma, A., et al. (2004). Structure-based thresholds of toxicological concern (TTC): Guidance for application to substances present at low levels in the diet. *Food and Chemical Toxicology*, *42*(1), pp. 65–83.

Kroes, R., Renwick, A. G., Feron, V., Galli, C. L., Gibney, M., Greim, H., et al. (2007). Application of the threshold of toxicological concern (TTC) to the safety evaluation of cosmetic ingredients. *Food and Chemical Toxicology*, *45*(12), pp. 2533–2562.

Lapenna, S., & Worth, A. (2011). *Analysis of the Cramer classification scheme for oral systemic toxicity - implications for its implementation in Toxtree*. Joint Research Centre.

Laufersweiler, M. C., Gadagbui, B., Baskerville-Abraham, I. M., Maier, A., Willis, A., Scialli, A. R., et al. (2012). Correlation of chemical structure with reproductive and developmental toxicity as it relates to the use of the threshold of toxicological concern. *Regulatory Toxicology and Pharmacology*, *62*(1), pp. 160–182.

Leeman, W. R., Krul, L., & Houben, G. F. (2013). Complex mixtures: Relevance of combined exposure to substances at low dose levels. *Food and Chemical Toxicology*, *58*, pp. 141–148.

van Leeuwen, C. J. ., & Vermeire, T. G. (2007). *Risk Assessment of Chemicals* (C. J. van Leeuwen & T. G. Vermeire, Eds.). Dordrecht: Springer Netherlands.

Liao, C., Sitzmann, M., Pugliese, A., & Nicklaus, M. C. (2011). Software and resources for computational medicinal chemistry. *Future medicinal chemistry*, *3*(8), pp. 1057–1085.

Madden, J. C. (2010). Chapter 21. Toxicokinetic Considerations in Predicting Toxicity, in: M. Cronin & J. Madden (Eds.), *In Silico Toxicology : Principles and Applications*, pp. 531–557. RSC Publishing.

Miller, E. C., & Miller, J. a. (1981). Searches for ultimate chemical carcinogens and their reactions with cellular macromolecules. *Cancer*, *47*(10), pp. 2327–45.

Mostrag-szlichtyng, A., & Worth, A. (2010). *Review of QSAR Models and Software Tools for predicting Biokinetic Properties*. Joint Research Centre.

Muller, A., Blaude, M., Ihlemann, C., Bjorge, C., Ohlsson, A., & Gebel, T. (2012). A regulatory approach to assess the potency of substances toxic to the reproduction. *Regulatory Toxicology and Pharmacology*, *63*(1), pp. 97–105.

Munro, I. C., Ford, R. A., Kennepohl, E., & Sprenger, J. G. (1996). Thresholds of Toxicological Concern Based on Structure - Activity Relationships. *Drug Metabolism Reviews*, *28*(1&2), pp. 209–217.

Munro, I. C., Hlywka, J. J., & Kennepohl, E. M. (2002). Risk assessment of packaging materials. *Food Additives and Contaminants*, *19*(S1), pp. 3–12.

Munro, I. C., Kennepohl, E., & Kroes, R. (1999). A procedure for the safety evaluation of flavouring substances. *Food and Chemical Toxicology*, *37*(2-3), pp. 207–232.

Munro, I. I. C. (1990). Safety Assessment Procedures for Indirect Food Additives: An Overview Report. *Regulatory Toxicology and Pharmacology*, *12*, pp. 2–12.

Netzeva, T. I., Worth, A., Aldenberg, T., Benigni, R., Cronin, M. T. D., Gramatica, P., et al. (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. *Alternatives to laboratory animals : ATLA*, *33*(2), pp. 155–73.

Neumann, H. G. (2005). *Monocyclic aromatic amino and nitro compounds: toxicity, genotoxicity and carcinogenicity, classification in acarcinogen category*.

OECD. (1983). *Test No. 415: One-Generation Reproduction Toxicity Study*. OECD Publishing.

OECD. (2001a). *Test No. 416: Two-Generation Reproduction Toxicity*. OECD Publishing.

OECD. (2001b). *Test No. 414: Prenatal Development Toxicity Study*. OECD Publishing.

OECD. (2011). *Test No. 443: Extended One-Generation Reproductive Toxicity Study*. OECD Publishing.

OECD. (2013). *Guidance document on developing and assessing Adverse Outcome Pathways*. Organisation for Economic Co-operation and Development.

Partosch, F., Mielke, H., Stahlmann, R., Kleuser, B., Barlow, S., & Gundert-Remy, U. (2014). Internal threshold of toxicological concern values: enabling route-to-route extrapolation. *Archives of Toxicology*, pp. 941–948.

Patlewicz, G., Jeliazkova, N., Safford, R. J., Worth, a P., & Aleksiev, B. (2008). An evaluation

of the implementation of the Cramer classification scheme in the Toxtree software. *SAR and QSAR in environmental research*, *19*(5-6), pp. 495–524.

Pinalli, R., Croera, C., Theobald, A., & Feigenbaum, A. (2011). Threshold of toxicological concern approach for the risk assessment of substances used for the manufacture of plastic food contact materials. *Trends in Food Science & Technology*, *22*(9), pp. 523–534.

Prasanna, S., & Doerksen, R. J. (2009). Topological polar surface area: a useful descriptor in 2D-QSAR. *Current medicinal chemistry*, *16*(1), pp. 21–41.

Price, P. S., Hollnagel, H. M., & Zabik, J. M. (2009). Characterizing the noncancer toxicity of mixtures using concepts from the TTC and quantitative models of uncertainty in mixture toxicity. *Risk Analysis*, *29*(11), pp. 1534–1548.

Puzyn, T., Leszczynski, J., & Cronin, M. T. (2010). *Recent Advances in QSAR Studies* (T. Puzyn, J. Leszczynski, & M. T. Cronin, Eds.). Dordrecht: Springer Netherlands.

van Ravenzwaay, B., Dammann, M., Buesen, R., Flick, B., & Schneider, S. (2011). The threshold of toxicological concern for prenatal developmental toxicity. *Regulatory Toxicology and Pharmacology*, *59*(1), pp. 81–90.

van Ravenzwaay, B., Dammann, M., Buesen, R., Flick, B., & Schneider, S. (2012). The threshold of toxicological concern for prenatal developmental toxicity in rabbits and a comparison to TTC values in rats. *Regulatory Toxicology and Pharmacology*, *64*(1), pp. 1–8.

Renwick, A. G. (2004). Risk characterisation of chemicals in food. *Toxicology Letters*, *149*(1-3), pp. 163–176.

Roberts, D. W., Api, A. M., Safford, R. J., & Lalko, J. F. (2015). Principles for identification of High Potency Category Chemicals for which the Dermal Sensitisation Threshold (DST) approach should not be applied. *Regulatory Toxicology and Pharmacology*, *72*(3), pp.

683–693.

Roberts, D. W., Aptula, A., Schultz, T. W., Shen, J., Api, A. M., Bhatia, S., et al. (2015). A practical guidance for Cramer class determination. *Regulatory Toxicology and Pharmacology*, *73*(3), pp. 971–984.

Rulis, A. M. (1986). De minimis and the Threshold of Regulation, in: C. W. Felix (Ed.), *Food Protection Technology*, pp. 29–37. Lewis Publishing, Inc.

Safford, R. J. (2008). The Dermal Sensitisation Threshold-A TTC approach for allergic contact dermatitis. *Regulatory Toxicology and Pharmacology*, *51*(2), pp. 195–200.

Safford, R. J., Api, A. M., Roberts, D. W., & Lalko, J. F. (2015). Extension of the Dermal Sensitisation Threshold (DST) approach to incorporate chemicals classified as reactive. *Regulatory Toxicology and Pharmacology*, *72*(3), pp. 694–701.

Safford, R. J., Aptula, A. O., & Gilmour, N. (2011). Refinement of the Dermal Sensitisation Threshold (DST) approach using a larger dataset and incorporating mechanistic chemistry domains. *Regulatory Toxicology and Pharmacology*, *60*(2), pp. 218–224.

Schnabel, J., & Taylor, S. (2015). Estimation of toxic hazard – A revised Cramer–Ford–Hall decision tree. *Toxicology Letters*, *238*(2), p. S7.

Schüürmann, G., Ebert, R. U., Tluczkiewicz, I., Escher, S. E., & Kühne, R. (2016). Inhalation threshold of toxicological concern (TTC) - Structural alerts discriminate high from low repeated-dose inhalation toxicity. *Environment International*, *88*, pp. 123–132.

Serafimova, R., Gatnik, M. F., & Worth, A. (2010). *Review of QSAR Models and Software Tools for Predicting Genotoxicity and Carcinogenicity*. Joint Research Centre.

Solecki, R., Kortenkamp, A., Bergman, Å., Chahoud, I., Degen, G., Dietrich, D., et al. (2016). *Scientific principles for the identification of endocrine disrupting chemicals – a consensus statement*. German Federal Institute for Risk Assessment (BfR).

Stanard, B., Dolan, D. G., Hanneman, W., Legare, M., & Bercu, J. P. (2015). Threshold of toxicological concern (TTC) for developmental and reproductive toxicity of anticancer compounds. *Regulatory Toxicology and Pharmacology*, *72*(3), pp. 602–609.

Tollefsen, K. E., Scholz, S., Cronin, M. T., Edwards, S. W., de Knecht, J., Crofton, K., et al. (2014). Applying Adverse Outcome Pathways (AOPs) to support Integrated Approaches to Testing and Assessment (IATA). *Regulatory Toxicology and Pharmacology*, *70*(3), pp. 629–640.

Tolls, J., Müller, M., Willing, A., & Steber, J. (2009). A New Concept for the Environmental Risk Assessment of Poorly Water Soluble Compounds and Its Application to Consumer Products. *Integrated Environmental Assessment and Management*, *5*(3), p. 374.

Vandenberg, L. N., Colborn, T., Hayes, T. B., Heindel, J. J., Jacobs, D. R., Lee, D.-H., et al. (2012). Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses. *Endocrine Reviews*, *33*(3), pp. 378–455.

Williams, E. S., Berninger, J. P., & Brooks, B. W. (2011). Application of chemical toxicity distributions to ecotoxicology data requirements under REACH. *Environmental Toxicology and Chemistry*, *30*(8), pp. 1943–1954.

Williams, F. M., Rothe, H., Barrett, G., Chiodini, A., Whyte, J., Cronin, M. T. D., et al. (2016). Assessing the safety of cosmetic chemicals: consideration of a flux decision tree to predict dermally delivered systemic dose for comparison with oral TTC (Threshold of Toxicological Concern). *Regulatory Toxicology and Pharmacology*, *76*, pp. 174–186.

de Wolf, W., Siebel-Sauer, A., Lecloux, A., Koch, V., Holt, M., Feijtel, T., et al. (2005). Mode of action and aquatic exposure thresholds of no concern. *Environmental toxicology and chemistry / SETAC*, *24*(2), pp. 479–85.

Worth, A., Bassan, A., Fabjan, E., Gallegos Saliner, A., Netzeva, T., Patlewicz, G., et al. (2007). *The Use of Computational Methods in the Grouping and Assessment of Chemicals -*

Preliminary Investigations. Joint Research Centre.

Worth, A., Cronin, M., Enoch, S., & Fioravanzo, E. (2012). *Applicability of the Threshold of Toxicological Concern (TTC) approach to cosmetics – preliminary analysis*. Joint Research Centre.

Worth, A., Lapenna, S., Piparo, E. Lo, & Serafimova, R. (2011). *A Framework for assessing in silico Toxicity Predictions : Case Studies with selected Pesticides*. Joint Research Centre.

Worth, A. P., & Cronin, M. T. D. (2001). The use of bootstrap resampling to assess the uncertainty of Cooper statistics. *ATLA Alternatives to Laboratory Animals*, *29*(4), pp. 447–459.

Yang, C., Arvidson, K., Cheeseman, M., Cronin, M. T. D., Enoch, S., Escher, S., et al. (2016). Development Of A Master Database Of Non-Cancer Threshold Of Toxicological Concern And Potency Categorization Based On ToxPrint Chemotypes, in: *SOT 2016: 55th Annual Meeting and ToxExpo*.

Yang, C., Cheeseman, M. A., & Worth, A. (2014). *The Concept of Threshold of Toxicological Concern and Recent Trends, In: SEUART-1 report*.

Zarn, J. a., Hänggi, E., & Engeli, B. E. (2015). Impact of study design and database parameters on NOAEL distributions used for toxicological concern (TTC) values. *Regulatory Toxicology and Pharmacology*, *72*, pp. 491–500.

9 Appendix

9.1 List of Nitrobenzenes with Ames Genotoxicity Results

Dataset was used for the analysis presented in the Chapter 2.

SMILES	CAS	Chemical name	Ames test result
c1cc(ccc1C)[N+](=O)[O-]	99-99-0	Toluene, p-nitro-	1
c1(c(c(ccc1)N=O)C)[N+](=O)[O-]	143922-95-6	Benzene, 2-methyl-1-nitro-3-nitroso-	0
O(C)c1c([N+](=O)[O-])cccc1	91-23-6	Benzene, 1-methoxy-2-nitro-	1
c1cc(ccc1N)[N+](=O)[O-]	100-01-6	4-Nitroaniline	1
c1cc(ccc1C(=O)O)[N+](=O)[O-]	62-23-7	Benzoic acid, *p*-nitro-	1
[Na+].S(=O)(=O)([O-])c1c([N+](=O)[O-])cc([N+](=O)[O-])cc1	885-62-1	Sodium 2,4-dinitrobenzenesulphonate	0
c1cccc(c1C(F)(F)F)[N+](=O)[O-]	384-22-5	Benzene, 1-nitro-2-(trifluoromethyl)-	0
c1(c(c(cc(c1)C(F)(F)F)[N+](=O)[O-])Cl)[N+](=O)[O-]	393-75-9	4-Chloro-3,5-dinitro-alpha,alpha,alpha-trifluorotoluene	0
c1cc(cc(c1O)[N+](=O)[O-])[N+](=O)[O-]	51-28-5	2,4-dinitrophenol	0
c1cc(c(cc1N)N)[N+](=O)[O-]	5131-58-8	*m*-Phenylenediamine, 4-nitro-	1
c1cc(cc(c1N)[N+](=O)[O-])N	5307-14-2	*p*-Phenylenediamine, 2-nitro-	1
c1cc(ccc1Cl)[N+](=O)[O-]	100-00-5	Benzene, 1-chloro-4-nitro-	1
c1c(ccc(c1)O)[N+](=O)[O-]	100-02-7	Phenol, *p*-nitro-	0
c1cc(ccc1[N+](=O)[O-])CCl	100-14-1	Benzene, 1-(chloromethyl)-4-nitro-	1
c1c(ccc(c1)NC)[N+](=O)[O-]	100-15-2	Benzenamine, *N*-methyl-4-nitro-	1
CC(=O)c1ccc([N+](=O)[O-])cc1	100-19-6	4'-nitroacetophenone	1
c1cc(ccc1CCO)[N+](=O)[O-]	100-27-6	Benzeneethanol, 4-nitro-	1
c1(ccc(cc1)CC)[N+](=O)[O-]	100-12-9	Benzene, 1-ethyl-4-nitro-	0
c1(c(c(cc(c1Cl)Cl)Cl)Cl)[N+](=O)[O-]	117-18-0	Benzene, 1,2,4,5-tetrachloro-3-nitro-	1
c1(c(ccc(c1)Cl)C)[N+](=O)[O-]	89-59-8	Benzene, 4-chloro-1-methyl-2-nitro-	0
c1(c(cc(cc1)N)[N+](=O)[O-])C	119-32-4	Benzenamine, 4-methyl-3-nitro-	1
c1(ccc(cc1[N+](=O)[O-])N)O	119-34-6	Phenol, 4-amino-2-nitro-	1
c1(c([N+](=O)[O-])cccc1)S(=O)(=O)O	80-82-0	*O*-nitrobenzenesulfonic Acid	0
[N+](=O)([O-])c1cc([N+](=O)[O-])ccc1C	121-14-2	2,4-Dinitrotoluene (containing 1.0-1.5% 2,6-dinitrotoluene)	1
c1(c(cc(cc1)C(F)(F)F)[N+](=O)[O-])Cl	121-17-5	Toluene, 4-chloro-.alpha.,.alpha.,.alpha.-trifluoro-3-nitro-	0
[As](=O)(c1cc([N+](=O)[O-])c(cc1)O)(O)O	121-19-7	Benzenearsonic acid, 4-hydroxy-3-nitro-	1
c1(cc(ccc1)Cl)[N+](=O)[O-]	121-73-3	*m*-Chloronitrobenzene	0
c1(ccc(c(c1)O)N)[N+](=O)[O-]	121-88-0	Phenol, 2-amino-5-nitro-	1
[N+](=O)(c1cccc(c1)C(=O)C)[O-]	121-89-1	Ethanone, 1-(3-nitrophenyl)-	1
[N+](=O)(c1cccc(c1)C(=O)Cl)[O-]	121-90-4	Benzoyl chloride, *m*-nitro-	1
c1(cccc(c1)C(=O)O)[N+](=O)[O-]	121-92-6	Benzoic acid, *m*-nitro-	1
[N+](=O)(c1ccc(cc1)C(=O)Cl)[O-]	122-04-3	*P*-nitrobenzoyl Chloride	1
CCc1ccccc1[N+](=O)[O-]	612-22-6	Benzene, 1-ethyl-2-nitro-	0
c1(c(c(ccc1[N+](=O)[O-])C)[N+](=O)[O-])S(=O)(=O)O	63348-71-0	Toluene-3-sulfonic acid, 2,4-dinitro-	0
c1ccc(c(c1[N+](=O)[O-])C(=O)O)C	13506-76-8	Benzoic acid, 2-methyl-6-nitro-	1

SMILES	CAS	Chemical name	Ames test result
c1(c(cc(cc1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])C(=O)O	129-66-8	2,4,6-trinitrobenzoic Acid	1
c1ccc(c(c1)CCO)[N+](=O)[O-]	15121-84-3	Benzeneethanol, 2-nitro-	1
C(=O)(c1c(cc(cc1[N+](=O)[O-])N)[N+](=O)[O-])O	114168-48-8	Benzoic acid, 4-amino-2,6-dinitro-	1
CCCN(c1c(cc(cc1[N+](=O)[O-])C(F)(F)F)[N+](=O)[O-])CCC	1582-09-8	*p*-Toluidine, .alpha.,.alpha.,.alpha.-trifluoro-2,6-dinitro-*N,N*-dipropyl	1
C(=O)(c1c(cc(cc1[N+](=O)[O-])[N+](=O)[O-])N)O	140380-55-8	Benzoic acid, 2-amino-4,6-dinitro-	1
c1(c(ccc(c1Cl)[N+](=O)[O-])Cl)Cl	17700-09-3	4-Nitro-1,2,3-trichlorobenzene	1
CCOc1c(cc(cc1)NC(=O)C)[N+](=O)[O-]	1777-84-0	*N*-(4-Ethoxy-3-nitrophenyl)acetamide	1
c1(cc(c(c(c1)Br)N)[N+](=O)[O-])[N+](=O)[O-]	1817-73-8	Aniline, 2-bromo-4,6-dinitro-	1
c1(c(cc(c(c1)C)[N+](=O)[O-])[N+](=O)[O-])S(=O)(=O)O	52146-86-8	2,4-dinitrotoluene-5-sulfonic Acid	0
c1c(cc(c(c1Cl)[N+](=O)[O-])Cl)Cl	18708-70-8	1,3,5-Trichloro-2-nitrobenzene	0
c1cc(c(c(c1)C(=O)O)C)[N+](=O)[O-]	1975-50-4	*o*-Toluic acid, 3-nitro-	1
c1(ccc(c(c1)C(=O)O)C)[N+](=O)[O-]	1975-52-6	*o*-Toluic acid, 5-nitro-	1
c1(c(c(cc(c1)[N+](=O)[O-])Cl)Cl)Cl	20098-48-0	Benzene, 1,2,3-trichloro-5-nitro-	1
c1(c(cc(cc1[N+](=O)[O-])N)[N+](=O)[O-])CO	226711-12-2	Benzyl alcohol, 4-amino-2,6-dinitro-	1
c1(ccc(cc1)N(CC)CC)[N+](=O)[O-]	2216-15-1	Benzenamine, *N,N*-diethyl-4-nitro-	1
[N+](=O)([O-])c1c(Cl)c(Cl)c(c(c1Cl)Cl)OC	2438-88-2	Benzene, 1,2,4,5-tetrachloro-3-methoxy-6-nitro-	0
C(=C\C=C\C=O)/c1ccc(cc1)[N+](=O)[O-]	2608-48-2	5-(4-nitrophenyl)-2,4-pentadienal	1
O=C(CNC(=O)c1ccc(cc1)[N+](=O)[O-])O	2645-07-0	Glycine, *N*-(4-nitrobenzoyl)-	1
c1(c(ccc(c1)N(CCO)CCO)NC)[N+](=O)[O-]	2784-94-3	Ethanol, 2,2'-[[4-(methylamino)-3-nitrophenyl]imino]bis-	1
c1c(cc(c(c1)NCCO)[N+](=O)[O-])N	04/01/2871	2-(4-Amino-2-nitroanilino)ethanol	1
S=P(Oc1ccc(cc1)[N+](=O)[O-])(OC)OC	298-00-0	Phosphorothioic acid, *O,O*-dimethyl *O*-(4-nitrophenyl) ester	1
c1(c(cc(cc1[N+](=O)[O-])[N+](=O)[O-])N)CO	226711-13-3	Benzyl alcohol, 2-amino-4,6-dinitro-	1
c1(cc(c(cc1)[N+](=O)[O-])C)C(=O)O	3113-71-1	Benzoic acid, 3-methyl-4-nitro-	1
c1(ccc(c(c1)C(=O)O)[N+](=O)[O-])C	3113-72-2	5-Methyl-2-nitrobenzoic acid	1
N(N)c1c(cccc1)[N+](=O)[O-].Cl	56413-75-3	(2-nitrophenyl)hydrazine hydrochloride	1
[N+](=O)([O-])c1c(Cl)c(Cl)ccc1	3209-22-1	Benzene, 1,2-dichloro-3-nitro-	1
c1c(cc(c(c1)NCCO)[N+](=O)[O-])N(CCO)CCO	33229-34-4	2,2'-((4-((2-Hydroxyethyl)amino)-3-nitrophenyl)imino)bisethanol	1
C(=O)(/C=C/c1c(cccc1)[N+](=O)[O-])O	1013-96-3	(E)-3-(2-nitrophenyl)acrylic acid	0
c1(cccc(c1[N+](=O)[O-])C(=O)O)C	5437-38-7	Benzoic acid, 3-methyl-2-nitro-	1
c1(c(c(c(cc1[N+](=O)[O-])O)N)O)C	185376-54-9	2,4-dihydroxylamino-6-nitrotoluene	1

SMILES	CAS	Chemical name	Ames test result
C(=O)(O)c1c([N+](=O)[O-])cccc1	552-16-9	Benzoic acid, o-nitro- (8CI)	1
C(=O)(CC)OCc1cc(ccc1)[N+](=O)[O-]	174794-16-2	*M*-nitrobenzyl Propionate	0
c1(cc(ccc1)O)[N+](=O)[O-]	554-84-7	3-Nitrophenol	1
O(P(=S)(OCC)OCC)c1ccc(cc1)[N+](=O)[O-]	56-38-2	Phosphorothioic acid, *O,O*-diethyl *O*-(4-nitrophenyl) ester	1
c1ccc(c(c1[N+](=O)[O-])N)C	570-24-1	Benzenamine, 2-methyl-6-nitro-	1
C(=O)(C)c1c([N+](=O)[O-])cccc1	577-59-3	Ethanone, 1-(2-nitrophenyl)-	0
c1(c(cc(cc1)C)N)[N+](=O)[O-]	578-46-1	5-Methyl-*o*-nitroaniline	0
c1(ccc(c(c1)OCCO)NCCO)[N+](=O)[O-]	59820-43-8	HC Yellow No. 4	1
Nc1cc(c(c(c1)C(C)C)N)[N+](=O)[O-]	155379-81-0	4-amino-3-nitro-5-isopropylaniline	0
C(=O)(CC)OCc1c(cccc1)[N+](=O)[O-]	132663-51-5	*O*-nitrobenzyl Propionate	0
C(=O)(CCC)OCc1c(cccc1)[N+](=O)[O-]	119979-38-3	*o*-Nitrobenzyl butyrate	0
c1cc(c(c(c1)N)C)[N+](=O)[O-]	603-83-8	*o*-Toluidine, 3-nitro-	1
C(=O)(CCC)OCc1cc(ccc1)[N+](=O)[O-]	119613-13-7	*M*-nitrobenzyl Butyrate	0
c1(cccc(c1C)[N+](=O)[O-])[N+](=O)[O-]	606-20-2	Toluene, 2,6-dinitro-	1
Nc1cc(c(c(c1)CC(C)O)N)[N+](=O)[O-]	104535-31-1	4-amino-3-nitro-5-beta-hydroxypropylaniline	0
Nc1cc(c(c(c1)CCO)N)[N+](=O)[O-]	104535-30-0	4-amino-3-nitro-5-beta-hydroxyethylaniline	1
c1ccc(c(c1)[N+](=O)[O-])C(=O)Cl	610-14-0	Benzoyl chloride, *o*-nitro-	1
c1ccc(c(c1)[N+](=O)[O-])C(=O)N	610-15-1	Benzamide, *o*-nitro-	1
c1ccc(c(c1)CC#N)[N+](=O)[O-]	610-66-2	Acetonitrile, (*o*-nitrophenyl)-	1
c1c(cc(c(c1)[N+](=O)[O-])Cl)Cl	611-06-3	1,3-Dichloro-4-nitrobenzene	1
c1ccc(c(c1)CCl)[N+](=O)[O-]	612-23-7	Toluene, .alpha.-chloro-*o*-nitro-	1
[N+](=O)([O-])c1cc(N)c(cc1)C(=O)O	619-17-0	Benzoic acid, 2-amino-4-nitro	1
c1(cccc(c1)CCl)[N+](=O)[O-]	619-23-8	3-Nitrobenzyl chloride	1
c1c(ccc(c1)[N+](=O)[O-])C(=O)N	619-80-7	Benzamide, *p*-nitro-	1
c1c(cc(c(c1Cl)O)N)[N+](=O)[O-].Cl	62625-14-3	Phenol, 2-amino-6-chloro-4-nitro-, hydrochloride	1
Nc1c(c(c(cc1C)N)[N+](=O)[O-])C	97629-64-6	4-amino-3-nitro-2,6-dimethylaniline	0
c1(c(cc(c(c1)O)N)Cl)[N+](=O)[O-]	02/07/6358	Phenol, 2-amino-4-chloro-5-nitro-	1
c1(c(cc(c(c1)[N+](=O)[O-])N)C(C)C)N	82856-97-1	1,4-Benzenediamine, 2-(1-methylethyl)-5-nitro-	1
c1(c(cc(cc1)N=O)[N+](=O)[O-])C	82414-03-7	4-nitroso-2-nitrotoluene	0
[N+](=O)(c1c(COC(=O)C)cccc1)[O-]	77376-01-3	*O*-nitrobenzyl Acetate	0
c1(cccc(c1)[N+](=O)[O-])I	645-00-1	Benzene, 1-iodo-3-nitro-	0
c1(cccc(c1)[N+](=O)[O-])C(=O)N	645-09-0	Benzamide, *m*-nitro-	1
c1c(cc(c(c1[N+](=O)[O-])C)[N+](=O)[O-])NO	59283-75-9	Benzenamine, *N*-hydroxy-4-methyl-3,5-dinitro-	1
c1cc(cc(c1F)[N+](=O)[O-])[N+](=O)[O-]	70-34-8	Benzene, 1-fluoro-2,4-dinitro-	1
C(C)(C)(C)c1c([N+](=O)[O-])c(C)c(c(c1[N+](=O)[O-])C)C(=O)C	81-14-1	Musk Ketone	0
c1ccc(c(c1C)[N+](=O)[O-])C	81-20-9	Benzene, 1,3-dimethyl-2-nitro-	1

SMILES	CAS	Chemical name	Ames test result
[N+](=O)([O-])c1c(c(cc(c1)[N+](=O)[O-])NO)C	59283-76-0	Benzenamine, *N*-hydroxy-2-methyl-3,5-dinitro-	1
c1(cc(c(c(c1O)[N+](=O)[O-])Cl)Cl)Cl	82-62-2	3,4,6-Trichloro-2-nitrophenol	0
[N+](=O)([O-])c1c(Cl)c(Cl)c(c(c1Cl)Cl)Cl	82-68-8	Pentachloronitrobenzene	0
c1(cccc(c1C)[N+](=O)[O-])C	83-41-0	*o*-Xylene, 3-nitro-	1
c1c(c(c(c(c1C(C)(C)C)OC)[N+](=O)[O-])C)[N+](=O)[O-]	83-66-9	Anisole, 6-tert-butyl-3-methyl-2,4-dinitro-	1
c1(c(cc(cc1[N+](=O)[O-])OC)C(C)(C)C)O	59282-34-7	2-tert-Butyl-4-methoxy-6-nitrophenol	0
c1(c(cc(cc1)[N+](=O)[O-])N)N.Cl	53209-19-1	4-Nitrobenzene-1,2-diamine monohydrochloride	1
c1(c(c(c(c(c1)[N+](=O)[O-])Cl)Cl)Cl)Cl	879-39-0	2,3,4,5-Tetrachloronitrobenzene	0
[N+](=O)([O-])c1c(C)cccc1	88-72-2	Benzene, 1-methyl-2-nitro-	0
[N+](=O)([O-])c1c(Cl)cccc1	88-73-3	Benzene, 1-chloro-2-nitro-	1
c1ccc(c(c1)N)[N+](=O)[O-]	88-74-4	Aniline, *o*-nitro-	0
c1ccc(c(c1)O)[N+](=O)[O-]	88-75-5	2-Nitrophenol	0
[N+](=O)([O-])c1c(Cl)c([N+](=O)[O-])cc(c1)[N+](=O)[O-]	88-88-0	Benzene, 2-chloro-1,3,5-trinitro-	1
c1(cc(cc(c1O)[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-]	88-89-1	Phenol, 2,4,6-trinitro-	1
C(=O)(C)OCc1cc(ccc1)[N+](=O)[O-]	21388-97-6	*M*-nitrobenzyl Acetate	0
c1(c(ccc(c1)N)[N+](=O)[O-])O	16292-86-7	2-Nitro-5-aminophenol	1
c1(ccc(c(c1)[N+](=O)[O-])C)C	89-58-7	*p*-Xylene, 2-nitro-	1
[N+](=O)([O-])c1c(N)ccc(c1)C	89-62-3	*p*-Toluidine, 2-nitro-	1
c1(c(ccc(c1)Cl)N)[N+](=O)[O-]	89-63-4	Benzenamine, 4-chloro-2-nitro-	1
c1(c(cc(c(c1)[N+](=O)[O-])Cl)Cl)Cl	89-69-0	5-Nitro-1,2,4-trichlorobenzene	1
c1(c(ccc(c1)C)[N+](=O)[O-])C	89-87-2	Benzene, 2,4-dimethyl-1-nitro-	1
c1(c(c(ccc1)N)[N+](=O)[O-])O	14703-71-0	2-nitro-3-aminophenol	0
c1c(cc(c(c1N)O)S(=O)(=O)O)[N+](=O)[O-]	96-67-3	Benzenesulfonic acid, 3-amino-2-hydroxy-5-nitro-	0
[N+](=O)([O-])c1cc([N+](=O)[O-])cc(c1O)N	96-91-3	Phenol, 2-amino-4,6-dinitro-	1
c1(c(ccc(c1)C(=O)O)C)[N+](=O)[O-]	96-98-0	3-Nitro-*p*-toluic acid	1
c1cc(cc(c1Cl)[N+](=O)[O-])[N+](=O)[O-]	97-00-7	1-Chloro-2,4-dinitrobenzene	1
[N+](=O)([O-])c1cc([N+](=O)[O-])ccc1N	97-02-9	2,4-Dinitroaniline	1
c1(cc(ccc1)[N+](=O)[O-])C(F)(F)F	98-46-4	Toluene, alpha,alpha,alpha-trifluoro-*m*-nitro-	0
c1ccccc1[N+](=O)[O-]	98-95-3	Benzene, nitro-	0
[N+](=O)([O-])c1cc(C)ccc1	99-08-1	Benzene, 1-methyl-3-nitro-	0
c1cc(cc(c1)N)[N+](=O)[O-]	99-09-2	Benzenamine, 3-nitro-	1
[N+](=O)([O-])c1cc(C)cc(c1)C	99-12-7	Benzene, 1,3-dimethyl-5-nitro-	1
[N+](=O)([O-])c1cc(Cl)c(c(c1)Cl)N	99-30-9	Aniline, 2,6-dichloro-4-nitro-	1
c1(c(ccc(c1)[N+](=O)[O-])C)C	99-51-4	Benzene, 1,2-dimethyl-4-nitro-	1
c1c(cc(c(c1)N)C)[N+](=O)[O-]	99-52-5	Benzenamine, 2-methyl-4-nitro-	1
c1(c(ccc(c1)[N+](=O)[O-])Cl)Cl	99-54-7	Benzene, 1,2-dichloro-4-nitro	1
[N+](=O)([O-])c1cc(N)c(cc1)C	99-55-8	5-Nitro-ortho-toluidine	1
[N+](=O)([O-])c1cc(N)c(cc1)N	99-56-9	*o*-Phenylenediamine, 4-nitro-	1

SMILES	CAS	Chemical name	Ames test result
[N+](=O)([O-])c1cc(N)c(cc1)O	99-57-0	Phenol, 2-amino-4-nitro-	1
[N+](=O)([O-])c1cc(N)c(cc1)OC	99-59-2	*o*-Anisidine, 5-nitro-	1
c1(c(ccc(c1)NC(=O)C(C)C)[N+](=O)[O-])C(F)(F)F	13311-84-7	*m*-propionotoluidide,alpha,alpha,alpha-trifluoro-2-methyl-4'-nitro	0
c1(cc(cc(c1O)C)[N+](=O)[O-])[N+](=O)[O-]	534-52-1	4,6-Dinitro-*o*-cresol	1
c1(c([N+](=O)[O-])cccc1C)[N+](=O)[O-]	602-01-7	Benzene, 1-methyl-2,3-dinitro-	1
c1(c(cc(cc1)C)[N+](=O)[O-])[N+](=O)[O-]	610-39-9	Benzene, 4-methyl-1,2-dinitro-	1
[C@@H]([C@@H](c1ccc(cc1)[N+](=O)[O-])O)(NC(=O)C(Cl)Cl)COC(=O)CCC(=O)[O-].[Na+]	982-57-0	Butanedioic acid, mono[(2R,3R)-2-[(dichloroacetyl)amino]-3-hydroxy-3-(4-nitrophenyl)propyl] ester, monosodium salt	0
c1(cc(ccc1)[N+](=O)[O-])[N+](=O)[O-]	99-65-0	Benzene, *m*-dinitro-	1
[NH4+].c1(c(cc(cc1[N+](=O)[O-])Cl)Cl)[O-]	6609-49-0	Phenol, 2,4-dichloro-6-nitro-, ammonium salt	0
[N+](=O)(c1c(CBr)cccc1)[O-]	3958-60-9	alpha-Bromo-2-nitrotoluene	0
O=[N+]([O-])c1cccc(CBr)c1	3958-57-4	*m*-Nitrobenzyl bromide	0
O=[N+]([O-])c1ccc(cc1)CO/C(=N\C(C)C)NC(C)C	02/11/2978	Carbamimidic acid, N,N'-bis(1-methylethyl)-, (4-nitrophenyl)methyl ester	1
CC(c1c(c(c(c(c1[N+](=O)[O-])C)[N+](=O)[O-])C)[N+](=O)[O-])(C)C	81-15-2	Benzene, 1-(1,1-dimethylethyl)-3,5-dimethyl-2,4,6-trinitro-	0
c1(cc(ccc1)[N+](=O)[O-])CC(=O)O	1877-73-2	*M*-nitrophenylacetic Acid	0
O=C(O)Cc1c([N+](=O)[O-])cccc1	3740-52-1	Acetic acid, (*o*-nitrophenyl)-	1
O=C(O)Cc1ccc([N+](=O)[O-])cc1	104-03-0	4-Nitrophenylacetic acid	1
Oc1c(cc(cc1)C[C@@H](C(=O)O)N)[N+](=O)[O-]	621-44-3	3-nitro-l-tyrosine	0
[N+](=O)(c1cc(OCC)ccc1)[O-]	621-52-3	3-Nitrophenetole	0
Oc1c(N)c([N+](=O)[O-])ccc1	603-85-0	3-Nitro-2-aminophenol	0
c1(c(cccc1[N+](=O)[O-])N)O	603-87-2	2-Nitro-6-aminophenol	0
c1(cc(ccc1)OC)[N+](=O)[O-]	555-03-3	3-Nitroanisole	0
c1(c(cc(cc1[N+](=O)[O-])C(F)(F)F)[N+](=O)[O-])O	393-77-1	*p*-Cresol, alpha,alpha,alpha-trifluoro-2,6-dinitro-	0
c1(c(c(cc(c1O)[N+](=O)[O-])[N+](=O)[O-])O)[N+](=O)[O-]	82-71-3	1,3-Benzenediol, 2,4,6-trinitro-	0
Nc1cc(Cl)ccc1[N+](=O)[O-]	1635-61-6	Aniline, 5-chloro-2-nitro-	1
c1(ccc(cc1)[N+](=O)[O-])S(=O)(=O)O	138-42-1	*p*-Nitrobenzenesulfonic acid	0
c1(cc(cc(c1N)Cl)[N+](=O)[O-])[N+](=O)[O-]	3531-19-9	Benzenamine, 2-chloro-4,6-dinitro-	1
c1(c(cc(cc1)[N+](=O)[O-])Cl)N	121-87-9	Benzenamine, 2-chloro-4-nitro-	1
c1(N)c(cc(cc1)[N+](=O)[O-])C#N	17420-30-3	Benzonitrile, 2-amino-5-nitro-	1
Nc1c([N+](=O)[O-])cccc1[N+](=O)[O-]	606-22-4	2,6-dinitroaniline	1
c1(cc(cc(c1)N)[N+](=O)[O-])[N+](=O)[O-]	618-87-1	3,5-Dinitroaniline	1
c1(cc(c(cc1C)C)[N+](=O)[O-])[N+](=O)[O-]	616-72-8	*m*-Xylene, 4,6-dinitro-	1
c1(c(cc(cc1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])N([N+](=O)[O-])C	479-45-8	*N*-Methyl-*N*-2,4,6-tetranitrobenzenamine	1
c1(cc(ccc1N)[N+](=O)[O-])OC	97-52-9	2-Methoxy-4-nitroaniline	1
[N+](=O)(c1ccc(cc1)OC)[O-]	100-17-4	Benzene, 1-methoxy-4-nitro-	1

SMILES	CAS	Chemical name	Ames test result
c1(c(cc(cc1[N+](=O)[O-])[N+](=O)[O-])[N+](=O)[O-])OC	606-35-9	2,4,6-trinitroanisole	1
c1(c([N+](=O)[O-])ccc(c1[N+](=O)[O-])C)C	603-02-1	2,4-Dinitro-*m*-xylene	1
[N+](=O)(c1ccc(/N=N/N(C)C)cc1)[O-]	7227-92-1	1-(4-nitrophenyl)-3,3-dimethyltriazine	1
c1(c(c([N+](=O)[O-])cc(c1C)[N+](=O)[O-])N)[N+](=O)[O-]	22603-58-3	2,4,6-Trinitro-*m*-toluidine	1
Nc1c(c(cc(c1)[N+](=O)[O-])[N+](=O)[O-])C	35572-78-2	Benzenamine, 3,5-dinitro-2-methyl-(9CI)	1
Nc1cc(c(c(c1)[N+](=O)[O-])C)[N+](=O)[O-]	19406-51-0	4-amino-2,6-dinitrotoluene	1
[N+](=O)(c1c(c(cc(c1)[N+](=O)[O-])[N+](=O)[O-])C)[O-]	118-96-7	Benzene, 2-methyl-1,3,5-trinitro-	1
c1(c(cc(cc1)[N+](=O)[O-])N=O)C	82414-02-6	Toluene, 4-nitro-2-nitroso-	0
c1cc(c([N+](=O)[O-])cc1[N+](=O)[O-])C=O	528-75-6	2,4-dinitrobenzaldehyde	1
c1cc(c(C=O)c(c1)[N+](=O)[O-])[N+](=O)[O-]	606-31-5	Benzaldehyde, 2,6-dinitro-	1
[N+](=O)([O-])c1cc(C=O)ccc1	99-61-6	Benzaldehyde, 3-nitro-	1
c1(c(cccc1)[N+](=O)[O-])C=O	552-89-6	*o*-Nitrobenzaldehyde	0
c1(ccc(cc1)C=O)[N+](=O)[O-]	555-16-8	Benzaldehyde, 4-nitro-	1
c1(ccc(cc1)CBr)[N+](=O)[O-]	100-11-8	alpha-Bromo-4-nitrotoluene	1
c1(c(ccc(c1)[N+](=O)[O-])SC#N)[N+](=O)[O-]	1594-56-5	2,4-Dinitrophenylthiocyanate	1
c1(ccc(cc1)C=C)[N+](=O)[O-]	100-13-0	Styrene, *p*-nitro-	1
c1(c(ccc(c1)[N+](=O)[O-])C)NO	95860-07-4	Benzenamine, N-hydroxy-2-methyl-5-nitro-	0
c1(cc(c(cc1)C)[N+](=O)[O-])NO	43192-03-6	4-hydroxylamino-2-nitrotoluene	0
[N+](=O)(c1c(/C=C/C(=O)O)cccc1)[O-]	612-41-9	2-nitrocinnamic Acid	0
[N+](=O)(c1cc(N=[N+]=[NH-])ccc1F)[O-]	28166-06-5	4-Azido-1-fluoro-2-nitrobenzene	1
c1(c(cccc1)Br)[N+](=O)[O-]	577-19-5	1-Bromo-2-nitrobenzene	0
c1(cc(ccc1)Br)[N+](=O)[O-]	585-79-5	*m*-Bromonitrobenzene	1
O=[N+]([O-])c1ccc(cc1)Br	586-78-7	*P*-bromonitrobenzene	1
ClCc1cc(cc(c1)[N+](=O)[O-])[N+](=O)[O-]	74367-78-5	alpha-Chloro-3,5-dinitrotoluene	1
C(=O)(/C(=C/c1cc(c(c(c1)[N+](=O)[O-])O)O)C#N)N(CC)CC	130929-57-6	(*E*)-alpha-Cyano-*N,N*-diethyl-3,4-dihydroxy-5-nitrocinnamamide	0
c1([N+](=O)[O-])cc(c(cc1N)Cl)N	26196-45-2	2,5-Diamino-4-chloronitrobenzene	1
c1(c(c(c(c1)[N+](=O)[O-])N)C)C)N	97629-28-2	1,4-Benzenediamine, 2,3-dimethyl-5-nitro-	1
c1(c(c(c(c1)C)N)[N+](=O)[O-])C)N	155379-83-2	1,4-Benzenediamine, 2,5-dimethyl-3-nitro-	1
c1(c(cc(c(c1)[N+](=O)[O-])N)F)N	134514-27-5	4-amino-3-nitro-6-fluoroaniline	1
c1(c(cc(c(c1)[N+](=O)[O-])N)OC)N	25917-90-2	1,4-Benzenediamine, 2-methoxy-5-nitro-	1
c1(c(c(cc(c1)[N+](=O)[O-])N)C)N	59229-75-3	4-Nitro-2,6-toluenediamine	1
c1c(cc(c(c1N)C)[N+](=O)[O-])N	6629-29-4	1,3-Benzenediamine, 4-methyl-5-nitro-	1
c1(c(cc(c(c1)[N+](=O)[O-])N)C)N	25917-89-9	1,4-Benzenediamine, 2-methyl-5-nitro-	1
[N+](=O)(c1c(ccc(c1)Cl)Cl)[O-]	89-61-2	2,5-dichloronitrobenzene	1
[O-][N+](=O)c1c(c([N+](=O)[O-])c(c(C)c1C)C)C(C)(C)C	145-39-1	1-tert-Butyl-3,4,5-trimethyl-2,6-dinitrobenzene	0

SMILES	CAS	Chemical name	Ames test result
[N+](=O)(c1c(cccc1)[N+](=O)[O-])[O-]	528-29-0	Benzene, 1,2-dinitro-	1
[N+](=O)(c1ccc(cc1)[N+](=O)[O-])[O-]	100-25-4	Benzene, *p*-dinitro-	1
[N+](=O)(c1c(cc(cc1)OP(=S)(OC)OC)C)[O-]	122-14-5	Phosphorothioic acid, *O,O*-dimethyl *O*-(4-nitro-*m*-tolyl) ester	0
c1(ccc(cc1)OCC)[N+](=O)[O-]	100-29-8	*p*-Nitrophenetole	1
c1ccc(c(c1[N+](=O)[O-])CO)[N+](=O)[O-]	96839-34-8	Benzenemethanol, 2,6-dinitro-	0
c1(ccc(cc1)[N+](=O)[O-])COC(=O)CC	30039-44-2	*p*-Nitrobenzyl propionate	1
c1(OP(=O)(OC)OC)cc(c(cc1)[N+](=O)[O-])C	2255-17-6	Fenitrooxon	1
P(=O)(Oc1ccc(cc1)[N+](=O)[O-])(OCC)OCC	311-45-5	Phosphoric acid, diethyl 4-nitrophenyl ester	1
O=[N+]([O-])c1c(N)c(N)ccc1	3694-52-8	1,2-Benzenediamine, 3-nitro-	1
c1(c(cc(cc1)O)[N+](=O)[O-])[N+](=O)[O-]	577-71-9	3,4-Dinitrophenol	1
c1(cc(ccc1)[N+](=O)[O-])S(=O)(=O)O	98-47-5	3-Nitrobenzenesulphonic acid	0
c1([N+](=O)[O-])ccc(cc1)S	1849-36-1	*p*-Nitrothiophenol	1
[N+](=O)(c1cc(cc(c1)[N+](=O)[O-])[N+](=O)[O-])[O-]	99-35-4	Benzene, 1,3,5-trinitro-	1
c1(c(c([N+](=O)[O-])ccc1)O)[N+](=O)[O-]	573-56-8	2,6-Dinitrophenol	0
c1(cc(c(cc1)[N+](=O)[O-])O)[N+](=O)[O-]	329-71-5	2,5-dinitrophenol	1
c1(c([N+](=O)[O-])cccc1O)[N+](=O)[O-]	66-56-8	Phenol, 2,3-dinitro-	1
c1(cc(ccc1Cl)C(=O)O)[N+](=O)[O-]	96-99-1	3-Nitro-4-chlorobenzoic acid	1
C(=O)(c1cc(c(cc1)N(C)C)[N+](=O)[O-])OCC(CCCC)CC	134682-95-4	4-N,N-Dimethylamino-3-nitrobenzoic acid, 2-ethylhexyl ester	1
c1(c(ccc(c1)[N+](=O)[O-])C(=O)O)[N+](=O)[O-]	610-30-0	2,4-Dinitrobenzoic acid	1
c1(c(ccc(c1)[N+](=O)[O-])[N+](=O)[O-])C(=O)O	610-28-6	2,5-Dinitrobenzoic acid	1
c1(c(ccc(c1)C(=O)O)[N+](=O)[O-])[N+](=O)[O-]	528-45-0	Benzoic acid, 3,4-dinitro-	1
c1(cc(cc(c1)[N+](=O)[O-])[N+](=O)[O-])C(=O)O	99-34-3	3,5-Dinitrobenzoic acid	1
c1(c(c(cc(c1)C#N)I)O)[N+](=O)[O-]	1689-89-0	Nitroxinil	0
c1(cc(ccc1)C#N)[N+](=O)[O-]	619-24-9	Benzonitrile, *m*-nitro-	1
c1(c(cccc1)[N+](=O)[O-])C#N	612-24-8	*O*-nitrobenzonitrile	0
c1(ccc(cc1)C#N)[N+](=O)[O-]	619-72-7	4-Nitrobenzonitrile	0
c1(c(cccc1)OCC)[N+](=O)[O-]	610-67-3	*o*-Nitrophenetole	1
c1(ccc(cc1)NO)[N+](=O)[O-]	16169-16-7	4-Nitrophenylhydroxylamine	1
c1(ccc(cc1)NN)[N+](=O)[O-]	100-16-3	*p*-Nitrophenylhydrazine	1
c1(c(ccc(c1)[N+](=O)[O-])NN)[N+](=O)[O-]	119-26-6	2,4-dinitrophenylhydrazine	1
[N+](=O)(c1cc([N+](=O)[O-])cc(c1C)O)[O-]	497-56-3	3,5-Dinitro-o-cresol	0
OC(=O)/C=C/c1ccc([N+](=O)[O-])cc1	619-89-6	4-Nitrocinnamic acid	1
O=C(O)/C=C/c1cccc([N+](=O)[O-])c1	555-68-0	2-Propenoic acid, 3-(3-nitrophenyl)-	1

SMILES	CAS	Chemical name	Ames test result
[N+](=O)(c1ccc(cc1)/C=C/C=O)[O-]	1734-79-8	2-Propenal, 3-(4-nitrophenyl)-	1
C(=O)(CCC)OCc1ccc(cc1)[N+](=O)[O-]	35300-54-0	Butanoic acid, (4-nitrophenyl)methyl ester	1
c1(ccc(cc1)COC(=O)C)[N+](=O)[O-]	619-90-9	Benzenemethanol, 4-nitro-, acetate (ester)	1
[O-][N+](=O)c1ccc(cc1)CO	619-73-8	Benzenemethanol, 4-nitro-	0
c1(c(cccc1)[N+](=O)[O-])CO	612-25-9	Benzyl alcohol, *o*-nitro-	0
[O-][N+](=O)c1cccc(CO)c1	619-25-0	Benzyl alcohol, *m*-nitro-	0
c1(c(ccc(c1)[N+](=O)[O-])CO)[N+](=O)[O-]	4836-66-2	Benzyl alcohol, 2,4-dinitro-	1
c1(c(ccc(c1)N)CO)[N+](=O)[O-]	22996-17-4	Benzyl alcohol, 4-amino-2-nitro-	1
OCc1c(cccc1[N+](=O)[O-])N	98451-51-5	Benzyl alcohol, 2-amino-6-nitro-	0
c1(cc(c(cc1)CO)N)[N+](=O)[O-]	78468-34-5	2-amino-4-nitrobenzyl Alcohol	1
c1(cc(cc(c1)[N+](=O)[O-])[N+](=O)[O-])C(=O)Cl	99-33-2	3,5-Dinitrobenzoyl chloride	1

9.2 COSMOS TTC Dataset

Dataset was used for the analysis presented in Chapter 4.

CAS	INCI Name	Cramer class	NOEL
98-86-2	ACETOPHENONE	Low (Class I)	0,005
67-63-0	ISOPROPYL ALCOHOL	Low (Class I)	0,018
127-47-9	RETINYL ACETATE	Low (Class I)	0,200
1195-32-0	P,ALPHA-DIMETHYLSTYRENE	Low (Class I)	0,200
111-30-8	GLUTARAL	Low (Class I)	0,212
112-27-6	TRIETHYLENE GLYCOL	Low (Class I)	0,500
123-86-4	BUTYL ACETATE	Low (Class I)	0,500
576-26-1	2,6-XYLENOL	Low (Class I)	0,200
75-18-3	DIMETHYL SULPHIDE	Low (Class I)	0,600
142-83-6	2,4-HEXADIENAL	Low (Class I)	0,743
111-11-5	METHYL CAPRYLATE	Low (Class I)	1,200
127-51-5	ALPHA-ISOMETHYL IONONE; ALPHA-ISOMETHYL LONONE	Low (Class I)	1,367
102-87-4	TRILAURYLAMINE	Low (Class I)	1,667
79-69-6	5-METHYL-ALPHA-IONONE	Low (Class I)	1,967
79-81-2	RETINYL PALMITATE	Low (Class I)	2,400
124-30-1	STEARAMINE	Low (Class I)	3,000
151-10-0	RESORCINOL DIMETHYL ETHER	Low (Class I)	3,333
1123-85-9	HYDRATROPIC ALCOHOL	Low (Class I)	3,333
87-20-7	ISOAMYL SALICYLATE	Low (Class I)	1,567
103-82-2	PHENYLACETIC ACID	Low (Class I)	5,000
50-81-7	ASCORBIC ACID; ASCORBIC ACID (L-)	High (Class III)	5,500
4395-92-0	ISOPROPYL PHENYLACETALDEHYDE	Low (Class I)	6,333
103-37-7	BENZYL BUTYRATE	Low (Class I)	7,000
915-67-3	CI 16185	Low (Class I)	7,500
1166-52-5	DODECYL GALLATE	Low (Class I)	3,333
68-26-8	RETINOL	Low (Class I)	10,000
8013-90-9	IONONE; IONONE (BETA-); MIXED IONONES; MIXED IONONES (BETA-)	Low (Class I)	3,333
25152-84-5	TRANS, TRANS-2,4-DECADIENAL	Low (Class I)	11,300

CAS	INCI Name	Cramer class	NOEL
100-42-5	STYRENE	Low (Class I)	12,000
105-53-3	DIETHYL MALONATE	Low (Class I)	12,000
106-72-9	DIMETHYL HEPTENAL; DIMETHYLHEPTENAL	Low (Class I)	13,333
7779-78-4	4-METHYL-1-PHENYL-2-PENTANOL	Low (Class I)	13,333
105-54-4	ETHYL BUTYRATE	Low (Class I)	4,800
50-00-0	FORMALDEHYDE	Low (Class I)	15,000
125-12-2	ISOBORNYL ACETATE	Low (Class I)	5,000
1504-74-1	METHOXYCINNAMAL; *O*-METHOXYCINNAMALDEHYDE	Low (Class I)	15,667
93-53-8	HYDRATROPIC ALDEHYDE	Low (Class I)	16,667
112-34-5	BUTOXYDIGLYCOL	Low (Class I)	16,667
106-44-5	P-CRESOL	Low (Class I)	16,667
111-70-6	HEPTANOL	Low (Class I)	16,667
111-66-0	OCTENE	Low (Class I)	16,667
106-24-1	GERANIOL	Low (Class I)	16,667
111-48-8	THIODIGLYCOL	Low (Class I)	16,667
621-82-9	CINNAMIC ACID	Low (Class I)	17,800
117-81-7	DIETHYLBEXYL PHTHALATE	Low (Class I)	18,000
110-43-0	METHYL AMYL KETONE	Intermediate (Class II)	6,667
77-99-6	TRIMETHYLOLPROPANE	Low (Class I)	22,333
1393-63-1	ANNATTO; CI 75120	Low (Class I)	23,000
100-37-8	DIETHYL ETHANOLAMINE	Low (Class I)	25,000
123-31-9	HYDROQUINONE	Low (Class I)	8,333
142-47-2	SODIUM GLUTAMATE	Low (Class I)	25,000
102-71-6	TRIETHANOLAMINE	Low (Class I)	26,667
103-36-6	ETHYL CINNAMATE	Low (Class I)	26,667
99-86-5	ALPHA-TERPINENE	Low (Class I)	30,000
108-46-3	RESORCINOL	Low (Class I)	10,667
77-90-7	ACETYL TRIBUTYL CITRATE	Low (Class I)	33,333
104-20-1	METHOXYPHENYLBUTANONE	Low (Class I)	39,667
617-48-1	MALIC ACID (DL-)	Low (Class I)	40,000
120-50-3	ISOBUTYL BENZOATE	Low (Class I)	42,000
107-35-7	TAURINE	Low (Class I)	50,000
105-67-9	2,4-XYLENOL	Low (Class I)	16,667
928-96-1	3-HEXENOL	Low (Class I)	50,000
108-39-4	*M*-CRESOL	Low (Class I)	16,667
74-79-3	ARGININE; ARGININE (L-)	Intermediate (Class II)	50,000
93-92-5	METHYLBENZYL ACETATE	Low (Class I)	16,667
111-17-1	THIODIPROPIONIC ACID	Low (Class I)	57,000
150-90-3	DISODIUM SUCCINATE	Low (Class I)	60,000
2519-30-4	BRILLIANT BLACK 1; CI 28440	Low (Class I)	60,000
119-36-8	METHYL SALICYLATE	Low (Class I)	72,000
659-70-1	ISOAMYL ISOVALERATE	Low (Class I)	73,000
101-39-3	METHYLCINNAMIC ALDEHYDE	Low (Class I)	73,667
97-53-0	EUGENOL	Low (Class I)	26,433
121-79-9	PROPYL GALLATE	Low (Class I)	86,000
2611-82-7	ACID RED 18; CI 16255	Low (Class I)	86,000
577-11-7	DIETHYLHEXYL SODIUM SULFOSUCCINATE	Low (Class I)	86,000
112-14-1	OCTYL ACETATE	Low (Class I)	100,000
5471-51-2	RASPBERRY KETONE	Low (Class I)	33,333
111-62-6	ETHYL OLEATE	Low (Class I)	100,000

CAS	INCI Name	Cramer class	NOEL
67-64-1	ACETONE	Low (Class I)	33,333
100-51-6	BENZYL ALCOHOL	Low (Class I)	100,000
57-11-4	STEARIC ACID	Low (Class I)	100,000
123-66-0	ETHYL HEXANOATE	Low (Class I)	100,000
814-80-2	CALCIUM LACTATE	Low (Class I)	100,000
107-21-1	GLYCOL	Low (Class I)	100,000
87-99-0	XYLITOL	Low (Class I)	100,000
141-43-5	ETHANOLAMINE; ETHANOLARNINE	Low (Class I)	106,667
111-76-2	BUTOXYETHANOL	Low (Class I)	109,475
25013-16-5	BHA	Low (Class I)	38,333
111-27-3	HEXYL ALCOHOL	Low (Class I)	123,333
104-55-2	CINNAMAL	Low (Class I)	41,667
71-36-3	N-BUTYL ALCOHOL	Low (Class I)	41,667
84-74-2	DIBUTYL PHTHALATE	Low (Class I)	41,667
104-76-7	ETHYLHEXANOL	Low (Class I)	41,667
1948-33-0	TBHQ	Low (Class I)	129,000
111-87-5	CAPRYLIC ALCOHOL	Low (Class I)	130,000
100-41-4	ETHYLBENZENE	Low (Class I)	45,333
94-13-3	PROPYLPARABEN	Low (Class I)	150,000
99-76-3	METHYLPARABEN	Low (Class I)	150,000
134-03-2	SODIUM ASCORBATE	High (Class III)	151,667
108-95-2	PHENOL	Low (Class I)	153,000
64-18-6	FORMIC ACID	Low (Class I)	160,000
102-20-5	PHENETHYL PHENYLACETATE	Low (Class I)	166,667
123-51-3	ISOAMYL ALCOHOL	Low (Class I)	166,667
80-26-2	A-TERPINYL ACETATE	Low (Class I)	166,667
123-11-5	ANISALDEHYDE	Low (Class I)	166,667
103-23-1	DIETHYLHEXYL ADIPATE	Low (Class I)	170,000
133-37-9	TARTARIC ACID; TARTARIC ACID (DL-)	Low (Class I)	181,000
87-69-4	TARTARIC ACID; TARTARIC ACID (L-)	Low (Class I)	181,000
78-96-6	ISOPROPANOLAMINE; ISOPROPANOLAMME	Low (Class I)	200,000
68921-42-6	ACID BLUE 9 ALUMINUM LAKE; BLUE 1 LAKE	Low (Class I)	200,000
19224-26-1	PROPYLENE GLYCOL DIBENZOATE	Low (Class I)	211,000
3567-69-9	ACID RED 14; CI 14720	Low (Class I)	83,333
1334-78-7	TOLUALDEHYDE ISOMERS	Low (Class I)	83,333
124-04-9	ADIPIC ACID	Low (Class I)	250,000
111-90-0	ETHOXYDIGLYCOL	Low (Class I)	83,333
16521-38-3	ACID BLUE 74 ALUMINUM LAKE	Low (Class I)	250,000
97-54-1	ISOEUGENOL; LSOEUGENOL	Low (Class I)	250,000
123-35-3	MYRCENE	Low (Class I)	250,000
138-86-3	D,L-LIMONENE; LIMONENE; LIMONENE (DL-)	Low (Class I)	250,000
5989-27-5	D-LIMONENE; LIMONENE (D-)	Low (Class I)	250,000
89-83-8	THYMOL	Low (Class I)	256,688
1934-21-0	ACID YELLOW 23; CI 19140; YELLOW 5	Low (Class I)	259,000
56-86-0	GLUTAMIC ACID; GLUTAMIC ACID (L-)	Low (Class I)	288,000
89-78-1	D,L-MENTHOL; MENTHOL	Low (Class I)	296,000
100-52-7	BENZALDEHYDE	Low (Class I)	100,000
111-77-3	METHOXYDIGLYCOL	Low (Class I)	300,000
4075-81-4	CALCIUM PROPIONATE	Low (Class I)	300,000
78-83-1	2-METHYLPROPANOL	Low (Class I)	105,333
513-86-0	ACETOIN	Low (Class I)	110,000

CAS	INCI Name	Cramer class	NOEL
106-35-4	BUTYL ETHYL KETONE	Intermediate (Class II)	333,333
78-35-3	LINALYL ISOBUTYRATE	Low (Class I)	333,333
106-32-1	ETHYL OCTANOATE	Low (Class I)	333,333
93-28-7	EUGENYL ACETATE	Low (Class I)	333,333
540-18-1	AMYL BUTYRATE	Low (Class I)	333,333
103-41-3	BENZYL CINNAMATE	Low (Class I)	333,333
60-29-7	ETHYL ETHER	High (Class III)	166,667
67-56-1	METHYL ALCOHOL	Low (Class I)	166,667
60-18-4	TYROSINE; TYROSINE (L-)	Intermediate (Class II)	500,000
79-09-4	PROPIONIC ACID	Low (Class I)	500,000
109-94-4	ETHYL FORMATE	High (Class III)	166,667
140-11-4	BENZYL ACETATE	Low (Class I)	510,000
96-48-0	BUTYROLACTONE	Low (Class I)	175,000
88-09-5	2-ETHYLBUTYRIC ACID	Low (Class I)	178,000
860-22-0	ACID BLUE 74; CI 73015	Low (Class I)	550,000
108-88-3	TOLUENE	Low (Class I)	196,667
614-33-5	TRIBENZOIN	Low (Class I)	201,333
72-18-4	VALINE; VALINE (L-)	Low (Class I)	617,667
109-43-3	DIBUTYL SEBACATE	Low (Class I)	625,000
5743-27-1	CALCIUM ASCORBATE	High (Class III)	666,667
2783-94-0	CI 15985; YELLOW 6	Low (Class I)	226,000
110-17-8	FUMARIC ACID	Low (Class I)	720,000
139-05-9	SODIUM CYCLAMATE	Low (Class I)	720,000
64-19-7	ACETIC ACID	Low (Class I)	242,000
12225-21-7	ACID YELLOW 23 ALUMINUM LAKE	Low (Class I)	750,000
60-12-8	PHENETHYL ALCOHOL	Low (Class I)	833,333
63-91-2	PHENYLALANINE; PHENYLALANINE (L-)	Intermediate (Class II)	835,000
144-62-7	OXALIC ACID	Low (Class I)	840,000
65-85-0	BENZOIC ACID	Low (Class I)	887,000
123-29-5	ETHYL PELARGONATE	Low (Class I)	295,667
5392-40-5	CITRAL	Low (Class I)	295,667
105-87-3	GERANYL ACETATE	Low (Class I)	295,667
106-30-9	ETHYL HEPTANOATE	Low (Class I)	295,667
106-27-4	ISOAMYL BUTYRATE	Low (Class I)	295,667
141-78-6	ETHYL ACETATE	Low (Class I)	300,000
94-26-8	BUTYLPARABEN	Low (Class I)	300,000
6381-77-7	SODIUM ERYTHORBATE	High (Class III)	900,000
532-32-1	SODIUM BENZOATE	Low (Class I)	1.000,000
71-23-8	PROPYL ALCOHOL	Low (Class I)	1.000,000
25956-17-6	CI 16035; CURRY RED; RED 40	Low (Class I)	1.001,000
50-70-4	SORBITOL	Low (Class I)	1.200,000
111-46-6	DIETHYLENE GLYCOL	Low (Class I)	1.250,000
137-40-6	SODIUM PROPIONATE	Low (Class I)	1.283,000
73-32-5	ISOLEUCINE; LSOLEUCINE (L-)	Low (Class I)	1.333,333
121-33-5	VANILLIN	Low (Class I)	1.441,000
120-47-8	ETHYLPARABEN	Low (Class I)	1.441,000
50-81-7	ASCORBIC ACID; ASCORBIC ACID (L-)	High (Class III)	1.458,000
585-86-4	LACTITOL	Low (Class I)	1.472,000
616-38-6	DIMETHYL CARBONATE	Low (Class I)	491,667
7549-37-3	CITRAL DIMETHYL ACETAL	Low (Class I)	1.500,000
112-05-0	PELARGONIC ACID	Low (Class I)	1.500,000
62-54-4	CALCIUM ACETATE	Low (Class I)	1.500,000

CAS	INCI Name	Cramer class	NOEL
69-65-8	MANNITOL	Low (Class I)	1.600,000
58-86-6	XYLOSE	Low (Class I)	1.666,667
2353-45-9	CI 42053; GREEN 3	Low (Class I)	1.716,000
24634-61-5	POTASSIUM SORBATE	Low (Class I)	591,333
503-74-2	ISOVALERIC ACID	Low (Class I)	718,000
84-66-2	DIETHYL PHTHALATE	Low (Class I)	2.218,000
57-55-6	PROPYLENE GLYCOL	Low (Class I)	833,333
50-99-7	GLUCOSE; GLUCOSE (D-)	Low (Class I)	2.500,000
302-72-7	ALANINE; ALANINE (DL-)	Low (Class I)	2.500,000
72-19-5	THREONINE; THREONINE (L-)	Low (Class I)	3.000,000
123-95-5	BUTYL STEARATE	Low (Class I)	3.100,000
69-79-4	MALTOSE	Low (Class I)	3.300,000
78-37-5	LINALYL CINNAMATE	Low (Class I)	3.333,333
56-81-5	GLYCERIN	Low (Class I)	3.442,000
3844-45-9	ACID BLUE 9; BLUE 1; CI 42090	Low (Class I)	3.502,000
110-44-1	SORBIC ACID	Low (Class I)	3.602,000
5793-94-2	CALCIUM STEAROYL LACTYLATE	Low (Class I)	1.478,333
25383-99-7	SODIUM STEAROYL LACTYLATE	Low (Class I)	1.478,333
64-17-5	ALCOHOL; ALCOHOL DENAT.	Low (Class I)	5.241,000
139-06-0	CALCIUM CYCLAMATE	Low (Class I)	7.203,000
112-80-1	OLEIC ACID	Low (Class I)	7.500,000
60-33-3	LINOLEIC ACID	Low (Class I)	12.628,000
17373-89-6	2-HEXYLIDENECYCLOPENTANONE	Intermediate (Class II)	1,123
25155-30-0	SODIUM DODECYLBENZENESULFONATE	Intermediate (Class II)	2,000
122-40-7	AMYL CINNAMAL	Intermediate (Class II)	2,000
28434-00-6	ALLETHRINS	Intermediate (Class II)	2,700
3391-86-4	AMYLVINYL CARBINOL	Intermediate (Class II)	4,850
584-79-2	ALLETHRINS	Intermediate (Class II)	6,000
79-41-4	METHACRYLIC ACID	Intermediate (Class II)	8,400
1406-66-2	TOCOPHEROL	Intermediate (Class II)	13,333
2705-87-5	ALLYL CYCLOHEXYLPROPIONATE	Intermediate (Class II)	16,667
2051-78-7	ALLYL BUTYRATE	Intermediate (Class II)	16,667
68-04-2	SODIUM CITRATE; SODIUM CITRATE (ANHYDROUS)	Low (Class I)	22,000
128-37-0	BHT	Intermediate (Class II)	25,000
108-94-1	CYCLOHEXANONE	Intermediate (Class II)	27,872
142-19-8	ALLYL HEPTANOATE	Intermediate (Class II)	16,533
2835-39-4	ALLYL ISOVALERATE	Intermediate (Class II)	20,667
136-77-6	HEXYLRESORCINOL	Intermediate (Class II)	20,833
122-99-6	PHENOXYETHANOL	Intermediate (Class II)	26,667

CAS	INCI Name	Cramer class	NOEL
705-86-2	DELTA-DECALACTONE	Intermediate (Class II)	83,333
713-95-1	DELTA-DODECALACTONE	Intermediate (Class II)	83,333
7235-40-7	BETA-CAROTENE; CI 40800; CI 75130	Low (Class I)	89,000
431-03-8	DIACETYL	Intermediate (Class II)	30,000
584-03-2	1,2-BUTANEDIOL	Intermediate (Class II)	100,000
99-49-0	CARVONE	Intermediate (Class II)	41,667
123-68-2	ALLYL CAPROATE	Intermediate (Class II)	41,667
94-86-0	ETHOXY-PROPENYLPHENOL	High (Class III)	166,667
77-92-9	CITRIC ACID	Low (Class I)	241,000
59-02-9	TOCOPHEROL; TOCOPHEROL (D-)	High (Class III)	103,333
104-61-0	GAMMA-NONALACTONE	Intermediate (Class II)	360,000
104-67-6	GAMMA-UNDECALACTONE	Intermediate (Class II)	360,000
514-78-3	CI 40850	Intermediate (Class II)	166,667
2244-16-8	CARVONE; D-CARVONE	Intermediate (Class II)	750,000
121-32-4	ETHYL VANILLIN	High (Class III)	1.441,000
78-92-2	SEC-BUTANOL	Intermediate (Class II)	1.644,000
107-88-0	BUTYLENE GLYCOL	Intermediate (Class II)	6.883,000
50-14-6	ERGOCALCIFEROL	High (Class III)	0,000
58-85-5	BIOTIN	High (Class III)	0,015
1197-01-9	TRIMETHYLBENZYL ALCOHOL	High (Class III)	0,207
52918-63-5	DELTAMETHRIN	High (Class III)	0,333
698-10-2	5-ETHYL-3-HYDROXY-4-METHYLFURAN-2-ONE	High (Class III)	0,433
59-30-3	FOLIC ACID	High (Class III)	0,500
27538-10-9	2-ETHYL-4-HYDROXY-5-METHYLFURAN-3(2H)-ONE	High (Class III)	0,507
15707-24-1	2,3-DIETHYLPYRAZINE	High (Class III)	0,583
7696-12-0	TETRAMETHRIN	High (Class III)	0,600
91-76-9	BENZOGUANAMINE	High (Class III)	0,633
58-56-0	PYRIDOXINE HCL	Low (Class I)	0,900
16409-43-1	ISOBUTENYL METHYLTETRAHYDROPYRAN	High (Class III)	0,935
77-71-4	DM HYDANTOIN	High (Class III)	1,000
107-41-5	HEXYLENE GLYCOL	High (Class III)	1,000
91-57-6	2-METHYLNAPHTHALENE	High (Class III)	1,000
111-12-6	METHYL 2-OCTYNOATE	High (Class III)	1,067
103-50-4	DIBENZYL ETHER	High (Class III)	1,110
91-62-3	6-METHYLQUINOLINE	High (Class III)	1,133
541-15-1	CARNITINE	Low (Class I)	1,300
94-36-0	BENZOYL PEROXIDE	High (Class III)	1,400
28588-74-1	2-METHYLFURAN-3-THIOL	High (Class III)	1,667
15707-23-0	2-ETHYL-3-METHYLPYRAZINE	High (Class III)	1,667
93-18-5	ETHYL BETA-NAPHTHYL ETHER	High (Class III)	1,667
24295-03-2	METHYL THIAZOLYL KETONE	High (Class III)	1,667

CAS	INCI Name	Cramer class	NOEL
95-16-9	BENZOTHIAZOLE	High (Class III)	1,700
93-16-3	METHYL ISOEUGENOL	Low (Class I)	2,000
26172-55-4	METHYLCHLOROISOTHIAZOLINONE	High (Class III)	2,000
3658-77-3	DIMETHYLHYDROXY FURANONE	High (Class III)	2,050
17369-59-4	PROPYLIDENE PHTHALIDE	High (Class III)	2,183
51-03-6	PIPERONYL BUTOXIDE	High (Class III)	2,700
22047-25-2	METHYL PYRAZINYL KETONE	High (Class III)	2,733
2444-46-4	HYDROXYMETHOXYBENZYL PELARGONAMIDE	High (Class III)	2,877
52645-53-1	PERMETHRIN	High (Class III)	3,000
93-15-2	METHYL EUGENOL	Low (Class I)	3,333
79-07-2	CHLOROACETAMIDE; CHLOROACETARNIDE	High (Class III)	3,333
1438-94-4	1-FURFURYL-1H-PYRROLE	High (Class III)	3,733
83-88-5	LACTOFLAVIN; RIBOFLAVIN	High (Class III)	4,000
10222-01-2	DIBROMOCYANOACETAMIDE	High (Class III)	4,433
989-38-8	BASIC RED 1	High (Class III)	5,000
546-80-5	ISOPROPYL-METHYLBICYCLOHEXANONE	High (Class III)	1,667
75-09-2	DICHLOROMETHANE	High (Class III)	5,000
14667-55-1	TRIMETHYLPYRAZINE	High (Class III)	6,000
108-45-2	*M*-PHENYLENEDIAMINE	High (Class III)	2,000
53956-04-0	AMMONIUM GLYCYRRHIZATE	High (Class III)	6,000
119-61-9	BENZOPHENONE	High (Class III)	6,445
13494-06-9	3,4-DIMETHYLCYCLOPENTANE-1,2-DIONE	High (Class III)	6,667
113-48-4	ETHYLHEXYLBICYCLOHEPTENE DICARBOXIMIDE	High (Class III)	7,500
67-03-8	THIAMINE HCL	Low (Class I)	7,500
58-95-7	TOCOPHERYL ACETATE	High (Class III)	8,000
120-40-1	LAURAMIDE DEA	High (Class III)	8,333
3777-69-3	AMYLFURAN	High (Class III)	8,667
148-79-8	THIABENDAZOLE	High (Class III)	10,000
58-08-2	CAFFEINE	High (Class III)	10,100
3734-67-6	CI 18050	High (Class III)	12,000
93-08-3	2-ACETONAPHTHONE	High (Class III)	12,300
119-53-9	BENZOIN	High (Class III)	4,167
100-47-0	BENZONITRILE	High (Class III)	12,500
137-09-7	2,4-DIAMINOPHENOL HCL	High (Class III)	12,500
85-91-6	METHYL N-METHYLANTHRANILATE	High (Class III)	5,000
108-91-8	CYCLOHEXYLAMINE	High (Class III)	15,000
55406-53-6	IODOPROPYNYL BUTYLCARBAMATE; LODOPROPYNYL BUTYLCARBAMATE	High (Class III)	15,000
3380-34-5	TRICLOSAN	High (Class III)	15,000
79-11-8	CHLOROACETIC ACID	High (Class III)	15,000
147-24-0	DIPHENHYDRAMINE HCL	High (Class III)	5,333
3208-40-0	2-PHENYLPROPYLTETRAHYDROFURAN	High (Class III)	16,193
520-45-6	DEHYDROACETIC ACID	High (Class III)	16,500
100-86-7	DIMETHYLBENZYL CARBINOL	High (Class III)	16,667
3008-43-3	METHYLCYCLOHEXANEDIONE	High (Class III)	16,667
128-44-9	SODIUM SACCHARIN	High (Class III)	18,000
1124-11-4	TETRAMETHYLPYRAZINE	High (Class III)	18,333
118-71-8	MALTOL	High (Class III)	6,667
288-32-4	IMIDAZOLE	High (Class III)	20,000

CAS	INCI Name	Cramer class	NOEL
55418-52-5	PIPERONYL ACETONE	High (Class III)	22,033
151-21-3	SODIUM LAURYL SULFATE	High (Class III)	22,333
90-43-7	O-PHENYLPHENOL	High (Class III)	25,000
134-62-3	DIETHYL TOLUAMIDE	High (Class III)	25,000
133-06-2	CAPTAN	High (Class III)	25,000
4418-26-2	SODIUM DEHYDROACETATE	High (Class III)	25,000
83-67-0	THEOBROMINE	High (Class III)	25,000
101-54-2	N-PHENYL-P-PHENYLENEDIAMINE	High (Class III)	26,124
165450-17-9	NEOTAME	High (Class III)	30,000
98-01-1	FURFURAL	High (Class III)	30,000
39711-79-0	ETHYL MENTHANE CARBOXAMIDE	High (Class III)	33,333
1634-04-4	T-BUTYL METHYL ETHER	High (Class III)	33,333
136-45-8	DIPROPYL PYRIDINEDICARBOXYLATE	High (Class III)	34,100
77-83-8	ETHYL METHYLPHENYLGLYCIDATE	High (Class III)	35,000
10453-86-8	RESMETHRIN	High (Class III)	40,000
108-31-6	MALEIC ANHYDRIDE	High (Class III)	13,333
5307-14-2	2-NITRO-P-PHENYLENEDIANIINE	High (Class III)	43,000
92-48-8	6-METHYL COUMARIN	High (Class III)	50,000
110-91-8	MORPHOLINE	High (Class III)	50,000
119515-38-7	HYDROXYETHYL ISOBUTYL PIPERIDINE CARBOXYLATE	High (Class III)	50,000
6369-59-1	TOLUENE-2,5-DIAMINE SULFATE	High (Class III)	55,000
2475-45-8	DISPERSE BLUE 1	High (Class III)	20,667
148-24-3	OXYQUINOLINE	High (Class III)	73,000
22839-47-0	ASPARTAME	High (Class III)	24,667
91-22-5	QUINOLINE	High (Class III)	80,000
121-88-0	2-AMINO-5-NITROPHENOL	High (Class III)	33,333
103-90-2	ACETAMINOPHEN	High (Class III)	110,000
59820-43-8	HC YELLOW NO. 4	High (Class III)	115,000
120-32-1	CHLOROPHENE	High (Class III)	40,000
137-66-6	ASCORBYL PALMITATE	High (Class III)	125,000
20018-09-1	DIIODOMETHYLTOLYLSULFONE	High (Class III)	126,000
91-64-5	COUMARIN	High (Class III)	50,000
134-20-3	METHYL ANTHRANILATE	High (Class III)	50,000
126-92-1	SODIUM ETHYLHEXYL SULFATE	High (Class III)	170,000
104-46-1	ANETHOLE; TRANS-ANETHOLE	Low (Class I)	172,000
458-37-7	CI 75300; CURCUMIN	High (Class III)	62,000
123-30-8	P-AMINOPHENOL; P-ARNINOPHENOL	High (Class III)	65,333
16423-68-0	ACID RED 51; CI 45430	High (Class III)	200,000
126-13-6	SUCROSE ACETATE ISOBUTYRATE	High (Class III)	200,000
4940-11-8	ETHYL HYDROXYPYRONE	High (Class III)	200,000
55589-62-3	POTASSIUM ACESULFAME	High (Class III)	266,000
77-93-0	TRIETHYL CITRATE	High (Class III)	284,000
56038-13-2	SUCRALOSE	High (Class III)	294,000
33229-34-4	HC BLUE NO. 2	High (Class III)	103,333
81-07-2	SACCHARIN	High (Class III)	360,000
120-57-0	HELIOTROPINE	High (Class III)	360,000
7647-01-0	HYDROCHLORIC ACID	High (Class III)	360,000
117-39-5	QUERCETIN	High (Class III)	400,000
7585-39-9	CYCLODEXTRIN	High (Class III)	133,333
464-49-3	CAMPHOR; D-CAMPHOR	High (Class III)	400,000
57-13-6	UREA	Low (Class I)	450,000
8002-43-5	LECITHIN; PHOSPHATIDYLCHOAE; PHOSPHATIDYLCHOLINE	High (Class III)	475,000
6373-74-6	ACID ORANGE 3	High (Class III)	166,667

CAS	INCI Name	Cramer class	NOEL
25085-02-3	ACRYLAMIDE/SODIUM ACRYLATE COPOLYMER	Intermediate (Class II)	500,000
1260-17-9	CI 75470	High (Class III)	500,000
1390-65-4	CARMINE; CI 75470	High (Class III)	500,000
90-80-2	GLUCONOLACTONE	High (Class III)	500,000
26446-38-8	SUCROSE PALMITATE	High (Class III)	720,000
89-65-6	ERYTHORBIC ACID	High (Class III)	784,000
122-59-8	PHENOXYACETIC ACID	High (Class III)	800,000
71-55-6	TRICHLOROETHANE	High (Class III)	1.000,000
52225-20-4	TOCOPHERYL ACETATE	High (Class III)	1.000,000
139-33-3	DISODIUM EDTA	High (Class III)	333,333
100-97-0	METHENAMINE	High (Class III)	1.180,000
89-25-8	PHENYL METHYL PYRAZOLONE	High (Class III)	1.297,000
57-50-1	SUCROSE	High (Class III)	1.600,000
25168-73-4	SUCROSE STEARATE	High (Class III)	666,667
9005-38-3	ALGIN	High (Class III)	2.000,000
527-07-1	SODIUM GLUCONATE	High (Class III)	2.500,000
1338-43-8	SORBITAN OLEATE	High (Class III)	2.500,000
63-42-3	LACTOSE	High (Class III)	3.333,333
1338-39-2	SORBITAN LAURATE; SORBITAN MONOLAURATE	High (Class III)	3.442,000
1338-41-6	SORBITAN STEARATE	High (Class III)	3.602,000
131-99-7	DISODIUM INOSINATE	High (Class III)	1.774,000
2809-21-4	ETIDRONIC ACID	High (Class III)	8,333
4691-65-0	DISODIUM INOSINATE	High (Class III)	1.441,000

9.3 List of 198 Biocides Active Substances

Dataset was used for the analysis presented in Chapter 5.

ID	common name	CAS
1	Formaldehyde	50-00-0
2	Piperonyl butoxide / PBO	51-03-6
3	Bronopol	52-51-7
4	Chlorocresol	59-50-7
5	Dichlorvos	62-73-7
6	Ethanol	64-17-5
7	Formic acid	64-18-6
8	Benzoic acid	65-85-0
9	Propan-2-ol	67-63-0
10	Salicylic acid	69-72-7
11	Propan-1-ol	71-23-8
12	Ethylene oxide	75-21-8
13	Citric acid	77-92-9
14	2-chloroacetamide	79-07-2
15	Bromoacetic acid	79-08-3
16	Glycollic acid	79-14-1
17	Peracetic acid	79-21-0
18	L-(+)-lactic acid	79-33-4
19	Warfarin	81-81-2
20	Rotenone	83-79-4
21	Symclosene	87-90-1
22	Biphenyl-2-ol	90-43-7
23	Dichlorophen	97-23-4
24	Cinnamic aldehyde	104-55-2

ID	common name	CAS
25	Geraniol	106-24-1
26	Glyoxal	107-22-2
27	Hexa-2,4-dienoic acid / Sorbic acid	110-44-1
28	Glutaraldehyde	111-30-8
29	Nonanoic acid	112-05-0
30	Methynonylketone	112-12-9
31	DCDMH	118-52-5
32	Chlorophene	120-32-1
33	2-Phenoxyethanol	122-99-6
34	Cetylpyridinium chloride	123-03-5
35	Octanoic acid	124-07-2
36	Sodium Cacodylate	124-65-2
37	Nitromethylidynetrimethanol	126-11-4
38	Tosylchloramide sodium - Chloramin T	127-65-1
39	Potassium dimethyldithiocarbamate	128-03-0
40	Sodium dimethyldithiocarbamate	128-04-1
41	Warfarin sodium	129-06-6
42	Sodium 2-biphenylate	132-27-4
43	Captan	133-06-2
44	Folpet	133-07-3
45	DEET	134-62-3
46	Thiram	137-26-8
47	Ziram	137-30-4
48	Potassium methyldithiocarbamate	137-41-7
49	Metam-sodium	137-42-8
50	Disodium cyanodithiocarbamate	138-93-2
51	1,3-bis(hydroxymethyl)urea	140-95-4
52	Nabam	142-59-6
53	Lauric acid	143-07-7
54	Thiabendazole	148-79-8
55	Naled	300-76-5
56	Diuron	330-54-1
57	Diazinon	333-41-5
58	Decanoic acid	334-48-5
59	Cyanamide	420-04-2
60	Dazomet	533-74-4
61	Tolylfluanid	731-27-1
62	Oxypyrion	822-89-9
63	Dimethoxane	828-00-2
64	Terbutryn	886-50-0
65	Dichlofluanid	1085-98-9
66	Copper thiocyanate	1111-67-7
67	d-Tetramethrin	1166-46-7
68	4,5-dichloro-3H-1,2-dithiol-3-one	1192-52-5
69	2-Butanone, peroxide	1338-23-4
70	Monolinuron	1746-81-2
71	2,4-dichlorobenzyl alcohol	1777-82-8
72	Chlorothalonil	1897-45-6
73	Fluometuron	2164-17-2
74	4-(2-nitrobutyl)morpholine	2224-44-4
75	Diamine (*N*-(3-aminopropyl)-*N*-dodecylpropane-1,3-diamine)	2372-82-9
76	DTBMA	2527-58-4
77	BIT	2634-33-5
78	MIT	2682-20-4
79	Troclosene sodium	2893-78-9
80	Sodium dichloroisocyanurate dihydrate	51580-86-0
81	Bis(trichloromethyl) sulphone	3064-70-8
82	Triclosan	3380-34-5

ID	common name	CAS
83	Reaction products of ethylene glycol with paraformaldehyde (EGForm)	3586-55-8
84	Chlorophacinone	3691-35-8
85	Dipyrithione	3696-28-4
86	Sodium 2,4,6-trichlorophenolate	3784-03-0
87	Sodium pyrithione	3811-73-2
88	CTAC	4080-31-3
89	HHT	4719-04-4
90	TMAD	5395-50-6
91	MBM	5625-90-1
92	Coumatetralyl	5836-29-3
93	Terbuthylazine	5915-41-3
94	(R)-*p*-mentha-1,8-diene	5989-27-5
95	Methylene dithiocyanate	6317-18-6
96	DMDMH	6440-58-0
97	(2-bromo-2-nitrovinyl)benzene	7166-19-0
98	DDAC	7173-51-5
99	Prometryn	7287-19-6
100	Calcium dihexa-2,4-dienoate	7492-55-9
101	PVP-iodine (see iodine)	25655-41-8
102	Tetramethrin	7696-12-0
103	EDHO	7747-35-5
104	Pyrethrins	8003-34-7
105	Pyrethroids	8003-34-7
106	Sodium hydrogen 2,2'methylenebis[4-chlorophenolate]	10187-52-7
107	DBNPA	10222-01-2
108	Carbendazim	10605-21-7
109	Zineb	12122-67-7
110	Zinc pyrithione	13463-41-7
111	Dodecylguanidine monohydrochloride	13590-97-1
112	Potassium 2-biphenylate	13707-65-8
113	(benzyloxy)methanol	14548-60-8
114	Copper pyrithione	14915-37-8
115	Chlorotoluron	15545-48-9
116	Sodium p-chloro-*m*-cresolate	15733-22-9
117	Chloralose	15879-93-3
118	Benzoxonium chloride	19379-90-9
119	*p*-[(diiodomethyl)sulphonyl]toluene	20018-09-1
120	TCMTB	21564-17-0
121	Bendiocarb	22781-23-3
122	Prallethrin	23031-36-9
123	Potassium Sorbate	24634-61-5
124	HPT	25254-50-6
125	OIT	26530-20-1
126	*Cis*-tricos-9-ene	27519-02-4
127	Dimethyloctadecyl[3-(trimethoxysilyl)propyl]ammonium chloride	27668-52-6
128	Cybutryne	28159-98-0
129	Bromadiolone	28772-56-7
130	ZE-TDA	31654-77-0
131	BCDMH / Bromochlorodimethylhydantoin	32718-18-6
132	3-(4-isopropylphenyl)-1,1-dimethylurea / Isoproturon	34123-59-6
133	Diflubenzuron	35367-38-5
134	Imazalil	35554-44-0
135	Azamethiphos	35575-96-3
136	DBDCB	35691-65-7
137	Cyphenothrin	39515-40-7
138	Dimethyltetradecyl[3-(trimethoxysilyl)propyl]ammonium chloride	41591-87-1
139	Citriodiol	42822-86-6
140	4,4-dimethyloxazolidine	51200-87-4

ID	common name	CAS
141	IR3535	52304-36-6
142	Cypermethrin	52315-07-8
143	Permethrin	52645-53-1
144	Deltamethrin	52918-63-5
145	Empenthrin	54406-48-3
146	IPBC	55406-53-6
147	Difenacoum	56073-07-5
148	Brodifacoum	56073-10-0
149	Propiconazole	60207-90-1
150	DCOIT	64359-81-5
151	Triflumuron	64628-44-0
152	Oxazolidin / MBO	66204-44-2
153	Cyromazine	66215-27-8
154	Fenpropimorph	67564-91-4
155	Cyfluthrin	68359-37-5
156	ADBAC (= BKC)	68424-85-1
157	Sodium *N*-(hydroxymethyl)glycinate	70161-44-3
158	Amines, C10-16-alkyldimethyl, *N*-oxides	70592-80-2
159	Octenidine dihydrochloride	70775-75-6
160	1,3-didecyl-2-methyl-1H-imidazolium chloride	70862-65-6
161	Fenoxycarb	72490-01-8
162	MMPP (MAGNESIUM MONOPEROXYPHTHALATE HEXAHYDRATE)	84665-66-7 (ex 14915-85-4)
163	TTPC	81741-28-8
164	Urea, *N,N'*-bis(hydroxymethyl)-, reaction products with 2-(2-butoxyethoxy)ethanol, ethylene glycol and formaldehyde	90604-54-9
165	Hexaflumuron	86479-06-3
166	DCEMH	89415-87-2
167	Tebuconazole	107534-96-3
168	Transfluthrin	118712-89-3
169	Etofenprox	80844-07-1
170	PAP	128275-31-0
171	Methylneodecanamide / MNDA	105726-67-8
172	Lambda cyhalothrin	91465-08-6
173	Flufenoxuron	101463-69-8
174	DCPP	3380-30-1
175	BBIT	4299-07-4
176	Flocoumafen	90035-08-8
177	Icaridine	119515-38-7
178	Fipronil	120068-37-3
179	cis CTAC	51229-78-8
180	Imidacloprid	138261-41-3
181	Thiamethoxam	153719-23-4
182	Imiprothrin	72963-72-5
183	Pyriproxyfen	95737-68-1
184	Bethoxazin	163269-30-5
185	Bis(3-aminopropyl)octylamine	86423-37-2
186	Clothianidin	210880-92-5
187	Peroxyoctanoic acid	33734-57-5
188	S-Methoprene	65733-16-6
189	Esfenvalerate	66230-04-4
190	alpha-Cypermethrin	67375-30-8
191	Cyproconazole	94361-06-5
192	Difethialone	104653-34-1
193	Chlorfenapyr	122453-73-0
194	Acetamiprid	160430-64-8
195	d-Phenothrin	188023-86-1
196	d-Allethrin	231937-89-6

ID	common name	CAS
197	Esbiothrin	260359-57-7
198	Spinosad	168316-95-8

9.4 Analysis of Structural Features for Biocides and Munro Datasets

Results of the analysis presented in Chapter 5.

Structure Set	No. of str. features in biocides	% of str. features in biocides	No. of str. features in Munro	% of str. features in Munro
alcohol	47	23,7	154	25,8
aldehyde	4	2,0	12	2,0
alkene	51	25,8	110	18,5
amines	66	33,3	179	30,0
carbamate	16	8,1	30	5,0
dithiocarbamate	10	5,1	4	0,7
dithiocarbamate(RSCSNR2)	1	0,5	1	0,2
dithiocarbamic acid (HSCSNHR)	5	2,5	0	
halide	63	31,8	174	29,2
bromide, alkenyl-	2	1,0	0	
bromide, alkenyl, acyc-	2	1,0	0	
bromide, alkyl-	5	2,5	7	1,2
bromide, aryl-	4	2,0	7	1,2
bromide, phenyl-	3	1,5	7	1,2
bromide, alkyl, acyc-	5	2,5	7	1,2
bromide, p-alkyl-	2	1,0	1	0,2
bromide, t-alkyl-	1	0,5	0	
chloride, alkenyl-	11	5,6	25	4,2
chloride, alkenyl, acyc-	9	4,5	16	2,7
chloride, alkenyl, cyc-	2	1,0	9	1,5
chloride, alkyl-	8	4,0	55	9,2
chloride, aryl-	31	15,7	81	13,6
chloride, alkyl, acyc-	8	4,0	46	7,7
chloride, p-alkyl-	1	0,5	24	4,0
chloride, phenyl-	25	12,6	74	12,4
misc oxygen groups	5	2,5	3	0,5
hydroperoxide	5	2,5	0	
hydroperoxide, alkyl	1	0,5	0	
peroxide	5	2,5	0	
peroxide, alkyl	1	0,5	0	
peroxyacid	4	2,0	0	
organometal	1	0,5	0	
urea	13	6,6	22	3,7
Heterocycles	77	38,9	185	31,0
imidazole	4	2,0	16	2,7
imidazolidine	7	3,5	8	1,3
isothiazole	1	0,5	0	
isothiazolidine	4	2,0	0	
oxazolidine	4	2,0	1	0,2

Structure Set	No. of str. features in biocides	% of str. features in biocides	No. of str. features in Munro	% of str. features in Munro
thiane(H)	1	0,5	0	
thiazole	4	2,0	2	0,3
1,3,5-triazine	5	2,5	15	2,5
1,3,5-triazine(H)	7	3,5	3	0,5
1,3,4-triazole	3	1,5	4	0,7
1,2,4-triazole	3	1,5	4	0,7

9.5 List of Biocides with NOEL

Dataset used for the analysis presented in Chapter 5.

CAS	common name	DB NAME	Study type	NOAEL
50-00-0	Formaldehyde	Munro	chr	15,00
51-03-6	Piperonyl butoxide / PBO	COSMOS	chr	2,70
62-73-7	Dichlorvos	Munro	chr	0,23
64-17-5	Ethanol	Rep Dose	sub	705,78
64-18-6	Formic acid	COSMOS	chr	160,00
65-85-0	Benzoic acid	Munro	repro	887,00
67-63-0	Propan-2-ol	Munro	terat	0,02
71-23-8	Propan-1-ol	COSMOS	Sub	1.000,00
77-92-9	Citric acid	COSMOS	terat	241,00
79-07-2	2-chloroacetamide	COSMOS	sub	3,33
79-08-3	Bromoacetic acid	Munro	terat	50,00
79-21-0	Peracetic acid	Rep Dose	sub	0,53
83-79-4	Rotenone	Toxref m	dev	0,25
90-43-7	Biphenyl-2-ol	COSMOS	dev	25,00
104-55-2	Cinnamic aldehyde	COSMOS	sub	41,67
106-24-1	Geraniol	COSMOS	sub	16,67
107-22-2	Glyoxal	Rep Dose	sub	42,33
110-44-1	Hexa-2,4-dienoic acid / Sorbic acid	COSMOS	chr	3.602,00
111-30-8	Glutaraldehyde	COSMOS	sub	0,21
112-05-0	Nonanoic acid	COSMOS	terat	1.500,00
120-32-1	Chlorophene	Rep Dose	sub	3,33
122-99-6	2-Phenoxyethanol	COSMOS	sub	26,67
128-04-1	Sodium dimethyldithiocarbamate	Toxref m	dev	1,67
133-06-2	Captan	Munro	terat	25,00
133-07-3	Folpet	Munro	terat	10,00
134-62-3	DEET	COSMOS	mgr	25,00
137-26-8	Thiram	Toxref m	dev	2,50
137-30-4	Ziram	Toxref m	dev	5,33
137-42-8	Metam-sodium	Toxref m	dev	6,67
142-59-6	Nabam	Toxref d	dev	2,50
148-79-8	Thiabendazole	Munro	terat	10,00
300-76-5	Naled	Munro	chr	0,20
330-54-1	Diuron	Munro	chr	1,00
333-41-5	Diazinon	Rep Dose	sub	0,03
420-04-2	Cyanamide	Toxref m	dev	3,33
533-74-4	Dazomet	Toxref m	dev	3,33
828-00-2	Dimethoxane	Munro	chr	62,50
886-50-0	Terbutryn	Munro	chr	0,10
1897-45-6	Chlorothalonil	Rep Dose	sub	0,50
2164-17-2	Fluometuron	Munro	sub	2,67

CAS	common name	DB NAME	Study type	NOAEL
3380-34-5	Triclosan	COSMOS	Chr	15,00
3691-35-8	Chlorophacinone	Toxref d	dev	0,004
5989-27-5	(R)-*p*-mentha-1,8-diene	Rep Dose	sub	1,67
6317-18-6	Methylene dithiocyanate	Toxref m	dev	2,00
7287-19-6	Prometryn	Munro	terat	12,00
7696-12-0	Tetramethrin	COSMOS	chr	0,60
10222-01-2	DBNPA	COSMOS	sub	4,43
10605-21-7	Carbendazim	Toxref d	dev	6,67
13463-41-7	Zinc pyrithione	COSMOS	dev	0,50
20018-09-1	*p*-[(diiodomethyl)sulphonyl]toluene	Munro	terat	126,00
21564-17-0	TCMTB	Toxref m	dev	25,50
22781-23-3	Bendiocarb	Toxref m	dev	3,33
23031-36-9	Prallethrin	Toxref m	dev	10,00
24634-61-5	Potassium Sorbate	COSMOS	sub	591,33
28772-56-7	Bromadiolone	Toxref m	dev	0,02
35367-38-5	Diflubenzuron	Munro	chr	2,00
35554-44-0	Imazalil	Munro	chr	10,00
35575-96-3	Azamethiphos	Toxref m	dev	66,67
52315-07-8	Cypermethrin	Munro	repro	0,50
52645-53-1	Permethrin	Munro	chr	3,00
52918-63-5	Deltamethrin	COSMOS	sub	0,33
55406-53-6	IPBC	COSMOS	mgr	15,00
60207-90-1	Propiconazole	Munro	chr	5,00
66215-27-8	Cyromazine	Munro	chr	1,50
68359-37-5	Cyfluthrin	Munro	repro	2,50
80844-07-1	Etofenprox	Toxref m	dev	4,17
105726-67-8	Methylneodecanamide / MNDA	Toxref m	dev	41,67
119515-38-7	Icaridine	COSMOS	dev	50,00
120068-37-3	Fipronil	Toxref m	dev	6,67
138261-41-3	Imidacloprid	Toxref m	dev	3,33
153719-23-4	Thiamethoxam	Toxref m	dev	66,67
95737-68-1	Pyriproxyfen	Toxref m	dev	100,00
210880-92-5	Clothianidin	Toxref m	dev	13,33
66230-04-4	Esfenvalerate	Toxref m	dev	0,83
94361-06-5	Cyproconazole	Toxref m	dev	4,00
122453-73-0	Chlorfenapyr	Toxref m	dev	25,00
168316-95-8	Spinosad	Toxref m	dev	66,67

9.6 List of Developmental Toxicants

Dataset used in the analysis presented in Chapter 6.

Substance name	CAS number	SMILES	Toxtree Cramer classification	EU[7]	NOAEL overall
nitrofen (ISO), 2,4-dichlorophenyl 4-nitrophenyl ether	1836-75-5	Clc2cc(Cl)ccc2Oc1ccc(N(=O)=O)cc1	High (Class III)	61	
dibutyltin dichloride	683-18-1	Cl[Sn](Cl)(CCCC)CCCC	High (Class III)	61	2,5

[7] EU Classification for reproductive toxicity

Substance name	CAS number	SMILES	Toxtree Cramer classification	EU[7]	NOAEL overall
ammonium pentadecafluorooctanoate	3825-26-1	FC(F)(C(F)(F)C(=O)O)C(F)(F)C(F)(F)C(F)(F)C(F)(F)C(F)(F)F	High (Class III)	61	1
2-Methoxyethanol	109-86-4	OCCOC	Low (Class I)	61	
Methoxyacetic acid	625-45-6	O=C(O)COC	High (Class III)	61	2,5
octabromobiphenyl ether	32536-52-0	Brc2c(Oc1cc(Br)c(Br)c(Br)c1Br)cc(Br)c(Br)c2Br	High (Class III)	61	2,5
dibutyltin di(acetate)	1067-33-0	O=C(C)O[Sn](CCCC)(OC(=O)C)CCCC	High (Class III)	61	5
Glufosinate ammonium	77182-82-2	CP(O)(=O)CCC(N)C(O)=O	High (Class III)	63	6,3
N-methylformamide	123-39-7	O=CNC	High (Class III)	61	10
carbendazim (ISO), methyl benzimidazol-2-ylcarbamate	10605-21-7	O=C(OC)Nc2nc1ccccc1n2	High (Class III)	61	10
Ethylene thiourea, ETU	96-45-7	S=C1NCCN1	High (Class III)	61	
2-(2-Aminoethylamino)ethanol	111-41-1	OCCNCCN	High (Class III)	61	10
Linuron (ISO) 3-(3,4-dichlorophenyl)-1-methoxy-1-methylurea	330-55-2	Clc1ccc(NC(=O)N(OC)C)cc1Cl	High (Class III)	61	12,5
tridemorph	24602-86-6	O1C(CN(CCCCCCCCCCCCCC)CC1C)C	High (Class III)	61	10
1,2-Diethoxyethane	629-14-1	CCOCCOCC	High (Class III)	61	25
1-2-Dimethoxyethane	110-71-4	COCCOC	High (Class III)	61	
bis(2-Methoxyethyl)ether	111-96-6	COCCOCCOC	High (Class III)	61	25
formamide	75-12-7	O=CN	High (Class III)	61	22,6
benomyl	17804-35-2	O=C(n1c2ccccc2nc1NC(=O)OC)NCCCC	High (Class III)	61	30
N,N-dimethylformamide, dimethyl formamide	68-12-2	O=CN(C)C	High (Class III)	61	44
Imidazole, *N*, *N*'-1,2-ethenediyl-methanimidamide	288-32-4	n1ccnc1	High (Class III)	61	60
2-Methoxypropanol	1589-47-5	OCC(OC)C	High (Class III)	61	63
2-Methoxypropyl_acetate	70657-70-4	O=C(OCC(OC)C)C	High (Class III)	61	86
dibutyl phthalate	84-74-2	O=C(OCCCC)c1ccccc1C(=O)OCCCC	Low (Class I)	61	
dinocap	39300-45-3	CCCCCCCCc1cc(c(OC(=O)\C=C/C)c(c1)N(=O)=O)N(=O)=O	High (Class III)	61	
1,2-bis(2-Methoxyethoxy)ethane	112-49-2	COCCOCCOCCOC	High (Class III)	61	125
flusilazole	85509-19-9	Fc1ccc(cc1)[Si](c2ccc(F)cc2)(C)Cn3ncnc3	High (Class III)	61	0,5
N-Methylcaprolactam	2556-73-2	O=C1N(C)CCCCC1	High (Class III)	61	30
diisopentylphthalate	605-50-5	CC(C)CCc1ccc(C(O)=O)c(C(O)=O)c1CCC(C)C	Low (Class I)	61	200

Substance name	CAS number	SMILES	Toxtree Cramer classification	EU[7]	NOAEL overall
N-methylacetamide	79-16-3	O=C(NC)C	High (Class III)	61	
1-methyl-2-pyrrolidone	872-50-4	O=C1N(C)CCC1	High (Class III)	61	125
1,2-benzenedicarboxylic acid; di-C6-8-branched alkylesters, C7-rich	71888-89-6	O=C(OCCCCC(C)C)c1ccccc1C(=O)OCCCCC(C)C	Low (Class I)	61	300
1,2-benzenedicarboxylic acid, di-C7-11-branched and linear alkylesters	68515-42-4	O=C(c1ccccc1C(=O)OCCCCCC(C)C)OCCCCCCCC	Low (Class I)	61	200
fluazifop-butyl	69806-50-4	O=C(OCCCC)C(Oc2ccc(Oc1ncc(cc1)C(F)(F)F)cc2)C	High (Class III)	61	1
butyl benzyl phthalate	85-68-7	O=C(OCCCC)c2ccccc2C(=O)OCc1ccccc1	Low (Class I)	61	50
Diisobutyl phthalate	84-69-5	O=C(OCC(C)C)c1ccccc1C(=O)OCC(C)C	Low (Class I)	61	250
tetrachloroethylene	127-18-4	Cl/C(Cl)=C(/Cl)Cl	High (Class III)	61	684
trichloroethylene, trichloroethene	79-01-6	Cl\C=C(/Cl)Cl	High (Class III)	63	
tetracarbonylnickel, nickel tetracarbonyl	13463-39-3	[Ni].[C-]#[O+].[C-]#[O+].[C-]#[O+].[C-]#[O+]	High (Class III)	61	
benzo[a]pyrene	50-32-8	c1ccc2c(c1)cc3ccc4cccc5c4c3c2cc5	High (Class III)	61	
Flumioxazin	103361-09-7	O=C3\C4=C(/C(=O)N3c2c(F)cc1OCC(=O)N(c1c2)CC#C)CCCC4	High (Class III)	61	10
Azafenidin	68049-83-2	Clc3c(OCC#C)cc(N1/N=C2\N(C1=O)CCCC2)c(Cl)c3	High (Class III)	61	2,03
2-Ethoxyethanol	110-80-5	CCOCCO	Low (Class I)	61	23
Vinclozolin	50471-44-8	O=C2OC(C(=O)N2c1cc(Cl)cc(Cl)c1)(\C=C)C	High (Class III)	61	5
N,N-dimethylacetamide	127-19-5	O=C(N(C)C)C	Low (Class I)	61	94
methyl mercury	22967-92-6	[Hg]C	High (Class III)	61	0,0015
bis(2-ethylhexyl) phthalate (DEHP)	117-81-7	O=C(OCC(CC)CCCC)c1ccccc1C(=O)OCC(CC)CCCC	Low (Class I)	61	4,8
Warfarin,	81-81-2,	CC(=O)CC(C\1=C(/O)c2ccccc2OC/1=O)c3ccccc3	High (Class III)	61	
dinoseb	88-85-7	CCC(C)c1cc(cc(c1O)N(=O)=O)N(=O)=O	High (Class III)	61	1
cycloheximide	66-81-9	O=C2NC(=O)CC(C[C@@H](O)[C@H]1C(=O)[C@@H](C)C[C@H](C)C1)C2	High (Class III)	61	
6-(2-chloroethyl)-6-(2-methoxyethoxy)-2,5,7,10-tetraoxa-6-silaundecane, etacelasil	37894-46-5	ClCC[Si](OCCOC)(OCCOC)OCCOC	High (Class III)	61	
2-ethylhexyl[[[3,5-bis(1,1-dimethylethyl)-4-	80387-97-9	O=C(OCC(CC)CCCC)CSCc1cc(c(O)c	High (Class III)	61	150

Substance name	CAS number	SMILES	Toxtree Cramer classification	EU[7]	NOAEL overall
hydroxyphenyl]methyl]thio] acetate		(c1)C(C)(C)C)C(C)(C)C			
methyl isocyanate	624-83-9	O=C=N/C	High (Class III)	63	
Fenarimol	60168-88-9	Clc1ccc(cc1)C(O)(c2ccccc2Cl)c3cncnc3	High (Class III)	63	2
Metconazole	125116-23-6	Clc1ccc(cc1)CC2C(O)(C(CC2)(C)C)Cn3ncnc3	High (Class III)	63	4
methyltin trichloride, trichloromethylstannane, MMTC	993-16-8	C[Sn](Cl)(Cl)Cl	High (Class III)	63	10
Epoxiconazole	133855-98-8	Fc1ccc(cc1)C3(OC3c2ccccc2Cl)Cn4ncnc4	High (Class III)	63	2,3
dimethyltin dichloride	753-73-1	Cl[Sn](Cl)(C)C	High (Class III)	63	10
ioxynil (ISO) and its salts, 4-hydroxy-3,5-diiodobenzonitrile	1689-83-4	Ic1cc(C#N)cc(I)c1O	High (Class III)	63	
myclobutanil, 2-(4-chlorophenyl)-2-(1H-1,2,4-triazol-1-ylmethyl)hexanenitrile	88671-89-0	Clc1ccc(cc1)C(C#N)(CCCC)Cn2ncnc2	High (Class III)	63	
bromoxynil (ISO) and its salts, 3,5-dibromo-4-hydroxybenzonitrile, bromoxynil phenol	1689-84-5	Brc1cc(C#N)cc(Br)c1O	High (Class III)	63	4
nonylphenol	25154-52-3	Oc1ccc(cc1)CCCCCCCCC	Intermediate (Class II)	63	10
cyproconazole (ISO), (2RS,3RS;2RS,3SR)-2-(4-chlorophenyl)-3-cyclopropyl-1-(1H-1,2,4-triazol-1-yl)butan-2-ol	94361-06-5	Clc1ccc(cc1)C(O)(C(C)C2CC2)Cn3ncnc3	High (Class III)	63	2
1-(4-chlorophenyl)-4,4-dimethyl-3-(1,2,4-triazol-1-ylmethyl)pentan-3-ol (tebuconazole)	107534-96-3	Clc1ccc(cc1)CCC(O)(C(C)(C)C)Cn2ncnc2	High (Class III)	63	10
Oxadiargyl	39807-15-3	O=C2O\C(=N/N2c1c(Cl)cc(Cl)c(OCC#C)c1)C(C)(C)C	High (Class III)	63	80
Maneb	12427-38-2	SC(=S)NCCNC(S)=S	High (Class III)	63	100
piperazine	110-85-0	C1CNCCN1	High (Class III)	63	94
chlorotoluron	15545-48-9	Clc1cc(NC(=O)N(C)C)ccc1C	High (Class III)	63	50
2-ethylhexanoic acid	149-57-5	O=C(O)C(CC)CCCC	Low (Class I)	63	
1-Bromopropane	106-94-5	BrCCC	High (Class III)	63	
1,3,5-trioxan, trioxymethylene	110-88-3	O1COCOC1	High (Class III)	63	100
2-(2-Methoxyethoxy)ethanol	111-77-3	OCCOCCOC	Low (Class I)	63	200
Isoxaflutole	141112-29-0	O=C(c1c(onc1)C2CC2)c3ccc(cc3S(=O)(=O)C)C(F)(F)F	High (Class III)	63	2
Quinoclamine	2797-51-5	O=C\2c1c(cccc1)C(=O)C(/Cl)=C/2N	High (Class III)	63	5
1,2,4-triazole	288-88-0	n1cnnc1	High (Class III)	63	

Substance name	CAS number	SMILES	Toxtree Cramer classification	EU[7]	NOAEL overall
thiourea, thiocarbamide	62-56-6	S=C(N)N	High (Class III)	63	
Fenpropimorph	67564-91-4	O2C(CN(CC(C)Cc1ccc(cc1)C(C)(C)C)CC2C)C	High (Class III)	63	15
amitrole (ISO), 1,2,4-triazol-3-ylamine	61-82-5	n1cnnc1N	High (Class III)	63	4
dodecachloropentacyclo[5.2.1.02,6.03,9.05,8]decane, mirex	2385-85-5	ClC53C1(Cl)C4(Cl)C2(Cl)C1(Cl)C(Cl)(Cl)C5(Cl)C2(Cl)C3(Cl)C4(Cl)Cl	High (Class III)	63	
C.I. Direct Blue 6, tetrasodium 3,3'-[[1,1'-biphenyl]-4,4'-diylbis(azo)]bis[5-amino-4-hydroxynaphthalene-2,7-disulphonate]	2602-46-2	c1cc(ccc1c2ccc(cc2)N=Nc3c(cc4cc(cc(c4c3O)N)S(=O)(=O)[O-])S(=O)(=O)[O-])N=Nc5c(cc6cc(cc(c6c5O)N)S(=O)(=O)[O-])S(=O)(=O)[O-].[Na+].[Na+].[Na+].[Na+]	High (Class III)	63	
C.I. Direct Red 28, disodium 3,3'-[[1,1'-biphenyl]-4,4'-diylbis(azo)]bis(4-aminonaphthalene-1-sulphonate)	573-58-0	OS(=O)(=O)c5cc(/N=N/c1ccc(cc1)c4ccc(/N=N/c3cc(c2ccccc2c3N)S(O)(=O)=O)cc4)c(N)c6c5cccc6	High (Class III)	63	
disodium 4-amino-3-[[4'-[(2,4-diaminophenyl)azo][1,1'-biphenyl]-4-yl]azo]-5-hydroxy-6-(phenylazo)naphtalene-2,7-disulphonate, C.I. Direct Black 38	1937-37-7	c1ccc(cc1)N=Nc2c(cc3cc(c(c(c3c2O)N)N=Nc4ccc(cc4)c5ccc(cc5)N=Nc6ccc(cc6N)N)S(=O)(=O)[O-])S(=O)(=O)[O-].[Na+].[Na+]	High (Class III)	63	
fentin hydroxide	76-87-9	O[Sn](c1ccccc1)(c1ccccc1)c1ccccc1	High (Class III)	63	0,3
Toluene	108-88-3	c1ccccc1C	Low (Class I)	63	660
propylenethiourea	2122-19-2	S=C1NCC(N1)C	High (Class III)	63	0,3

9.7 List of Reproductive Toxicants

Dataset was used for the analysis presented in Chapter 6.

Substance name	CAS number	SMILES	Toxtree Cramer classification	EU[8]	NOAEL overall
Phoxim	14816-18-3	N#C/C(=N\OP(=S)(OCC)OCC)c1ccccc1	High (Class III)	62	

[8] EU Classification for fertility

Substance name	CAS number	SMILES	Toxtree Cramer classification	EU[8]	NOAEL overall
1,2-dimethoxyethane, ethylene glycol dimethyl ether, EGDME	110-71-4	COCCOC	High (Class III)	60	
bis(2-methoxyethyl) ether	111-96-6	COCCOCCOC	High (Class III)	60	48
bis(2-ethylhexyl) phthalate, di-(2-ethylhexyl) phthalate, DEHP	117-81-7	O=C(OCC(CC)CCCC)c1ccccc1C(=O)OCC(CC)CCCC	Low (Class I)	60	37
1,2-benzenedicarboxylic acid, dipentylester, branched and linear [1] n-pentyl-isopentylphthalate [2] di-n-pentyl phthalate [3] diisopentylphthalate [4]	131-18-0 [3] 605-50-5 [4]	O=C(OCCCCC)c1ccccc1C(=O)OCCCCC/	Low (Class I)	60	
2,4-dinitrotoluene,	121-14-2 [1] 25321-14-6 [2]	Cc1ccc(cc1[N](=O)(=O))N(=O)(=O)	High (Class III)	62	
1,3-diphenylguanidine	102-06-7	C1=CC=C(C=C1)NC(=NC2=CC=CC=C2)N	High (Class III)	62	
acrylamide, prop-2-enamide	79-06-1	O=C(\C=C)N	Intermediate (Class II)	62	8
Linuron	330-55-2	Clc1ccc(NC(=O)N(OC)C)cc1Cl	High (Class III)	62	20
Benfuracarb	82560-54-1	O=C(OCC)CCN(SN(C(=O)Oc2cccc1c2OC(C1)(C)C)C)C(C)C	High (Class III)	62	1,9
octamethylcyclotetrasiloxane	556-67-2	O1[Si](O[Si](O[Si](O[Si]1(C)C)(C)C)(C)C)(C)C	High (Class III)	62	220
tris(2-chloroethyl) phosphate	115-96-8	ClCCOP(=O)(OCCCl)OCCCl	High (Class III)	60	175
Glufosinate ammonium	77182-82-2	CP(O)(=O)CCC(N)C(O)=O	High (Class III)	60	43,7
dibutyltin dichloride	683-18-1	Cl[Sn](Cl)(CCCC)CCCC	High (Class III)	60	3,8
nonylphenol [1]; 4-nonylphenol, branched [2]	25154-52-3 [1]; 84852-15-3 [2]	Oc1ccc(cc1)CCCCCCCCC	Intermediate (Class II)	62	10
1-Bromopropane	106-94-5	BrCCC	High (Class III)	60	
1,2-Dibromo-3-chloropropane	96-12-8	BrC(CBr)CCl	High (Class III)	60	
1,2,3-trichloropropane	96-18-4	ClCC(Cl)CCl	High (Class III)	60	30
Mirex	2385-85-5	ClC53C1(Cl)C4(Cl)C2(Cl)C1(Cl)C(Cl)(Cl)C5(Cl)C2(Cl)C3(Cl)C4(Cl)Cl	High (Class III)	62	
α,α,α,4-terachlorotoluene, p-chlorobenzotrichloride	5216-25-1	ClC(Cl)(Cl)c1ccc(Cl)cc1	High (Class III)	62	1,25
diphenylether; octabromo derivative, octabromobiphenyl ether, octabromobiphenyl ether derivative, OBDPO, OBDPE	32536-52-0	Brc2c(Oc1cc(Br)c(Br)c(Br)c1Br)cc(Br)c(Br)c2Br	High (Class III)	62	4,6
2-methoxyethanol	109-86-4	OCCOC	Low (Class I)	60	100
2-Ethoxyethanol	110-80-5	CCOCCO	Low (Class I)	60	93

Substance name	CAS number	SMILES	Toxtree Cramer classification	EU[8]	NOAEL overall
allyl glycidyl ether	106-92-3	C=CCOCC1CO1	High (Class III)	62	
Glycidol	556-52-5	OCC1OC1	High (Class III)	60	
Fenarimol	60168-88-9	Clc1ccc(cc1)C(O)(c2ccccc2Cl)c3cncnc3	High (Class III)	62	1
TEGDME	112-49-2	COCCOCCOCCOC	High (Class III)	62	635
2-(2-aminoethylamino)ethanol (AEEA)	111-41-1	OCCNCCN	High (Class III)	62	250
2,3-Epoxipropyltrimethylammonium chloride (EPTAC)	3033-77-0	C1(CN(C)(C)C)CO1	High (Class III)	62	10
Bisphenol A	80-05-7	Oc1ccc(cc1)C(c2ccc(O)cc2)(C)C	High (Class III)	62	50
phenolphthalein	77-09-8	O=C1OC(c2ccccc12)(c3ccc(O)cc3)c4ccc(O)cc4	High (Class III)	62	150
hexan-2-one, methyl butyl ketone, butyl methyl ketone, methyl-n-butyl ketone	591-78-6	O=C(C)CCCC	Intermediate (Class II)	62	
quinomethionate	2439-01-2	O=C1Sc2nc3ccc(cc3nc2S1)C	High (Class III)	62	3
Tepraloxydim	149979-41-9	CC/C(=N\OC\C=C\Cl)/C=1C(=O)CC(CC=1O)C2CCOCC2	High (Class III)	62	12
Vinclozolin	50471-44-8	O=C2OC(C(=O)N2c1cc(Cl)cc(Cl)c1)(\C=C)C	High (Class III)	60	5
Dibutyl phthalate	84-74-2	O=C(OCCCC)c1ccccc1C(=O)OCCCC	Low (Class I)	62	
Quizalofop-*P*-tefuryl	119738-06-6	O=C(OCC1CCCO1)[C@@H](C)Oc4ccc(Oc2cnc3cc(Cl)ccc3n2)cc4	High (Class III)	62	21
Benzylbutyl phthalate	85-68-7	O=C(OCCCC)c2ccccc2C(=O)OCc1ccccc1	Low (Class I)	62	100
Diisobutyl phthalate	84-69-5	O=C(OCC(C)C)c1ccccc1C(=O)OCC(C)C	Low (Class I)	62	
nitrobenzene	98-95-3	c1(N(=O)=O)ccccc1	High (Class III)	62	12
Dinoseb	88-85-7	CCC(C)c1cc(cc(c1O)N(=O)=O)N(=O)=O	High (Class III)	62	3,8
Azafenidin	68049-83-2	Clc3c(OCC#C)cc(N1/N=C2\N(C1=O)CCCC2)c(Cl)c3	High (Class III)	62	2,31
Cyclohexylamine	108-91-8	NC1CCCCC1	High (Class III)	62	30
2,4-toluenediamine 4-methyl-*m*-phenylenediamine	95-80-7	Nc1cc(N)c(cc1)C	High (Class III)	62	5
4,4'-oxydianiline and its salts, p-aminophenyl ether	101-80-4	O(c1ccc(N)cc1)c2ccc(cc2)N	High (Class III)	62	16
piperazine hydrochloride; [1] piperazine dihydrochloride; [2] piperazine phosphate [3]	6094-40-2 [1]	N1CCNCC1	High (Class III)	62	125

Substance name	CAS number	SMILES	Toxtree Cramer classification	EU[8]	NOAEL overall
Carbendazim	10605-21-7	O=C(OC)Nc2nc1ccccc1n2	High (Class III)	60	50
Benomyl	17804-35-2	O=C(n1c2ccccc2nc1NC(=O)OC)NCCCC	High (Class III)	60	30
Molinate	2212-67-1	O=C(SCC)N1CCCCCC1	High (Class III)	62	0,3
Epoxiconazole	133855-98-8	Fc1ccc(cc1)C3(OC3c2ccccc2Cl)Cn4ncnc4	High (Class III)	62	1,9
tetrahydro-1,3-dimethyl-1H-pyrimidin-2-one, dimethylpropyleneurea, DMPU, 1,3-dimethyltetrahydro-2(1H)-pyrimidone	7226-23-5	O=C1N(C)CCCN1C	High (Class III)	62	10
Ketoconazole	65277-42-1	O=C(N5CCN(c4ccc(OCC1OC(OC1)(c2ccc(Cl)cc2Cl)Cn3cncc3)cc4)CC5)C	High (Class III)	60	4
4-tert-butylbenzoic acid	98-73-7	O=C(O)c1ccc(cc1)C(C)(C)C	Low (Class I)	60	1,6
Methoxyacetic acid	625-45-6	O=C(O)COC	High (Class III)	60	