The potential of computer-based quantitative structure activity approaches for predicting acute toxicity of chemicals

Elton Zvinavashe

Promotoren: Prof. Dr. Ir. I.M.C.M. Rietjens
Hoogleraar in de Toxicologie
Wageningen Universiteit

Prof. Dr. A.J. Murk
Persoonlijk hoogleraar bij de leerstoelgroep
Toxicologie
Wageningen Universiteit

Co-promotor: Dr. A.P. Freidig
Product Manager
TNO Quality of Life, Zeist

Promotiecommissie: Prof. Dr. F.X.R. van Leeuwen
RIVM, Bilthoven

Prof. Dr. B.J. Blaauboer
Universiteit Utrecht

Prof. Dr. C. Hendriksen
Nederlands Vaccin Instituut (NVI), Bilthoven

Prof. Dr. A.A. Koelmans
Wageningen Universiteit

Dit onderzoek is uitgevoerd binnen de onderzoekschool WIMEK/SENSE

The potential of computer-based quantitative structure activity approaches for predicting acute toxicity of chemicals

Elton Zvinavashe

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit
prof. dr. M.J. Kropff
in het openbaar te verdedigen
op dinsdag 16 December 2008
des namiddags te vier uur in de Aula

Title: The potential of computer-based quantitative structure activity approaches for predicting acute toxicity of chemicals.

Author: Elton Zvinavashe
Thesis Wageningen University, Wageningen, The Netherlands (2008)
With abstract, with references, with summary in Dutch

ISBN: 978-90-8585-253-7

Dedicated to *my daughter, Heather Anesu*

Abstract

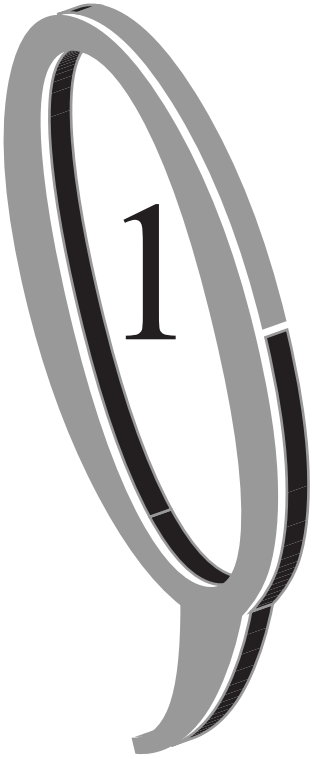
Within the EU, the management of the risks of chemicals currently falls under a new legislation called Registration, Evaluation, and Authorization of Chemicals (REACH). Within the next 10 years, the existing (eco)toxicological data gaps for the more than 100 000 chemicals on the European Inventory of Existing Commercial Substances (EINECS) should be filled. The challenge is to provide this toxicity information in a fast, cost effective manner, avoiding the use of experimental animals as much as possible. In this regard, REACH has provisions to allow for the use of *in vitro* and/or *in silico* methods, e.g. those based on (Quantitative) Structure Activity Relationships [(Q)SARs], to provide toxicity information or identify hazards of chemicals. This information can subsequently be used to identify priority chemicals for further risk evaluation. A QSAR is based on the assumption that the biological activity of a new or untested chemical can be inferred from the molecular structure, or properties of similar compounds whose activities have already been assessed. Therefore, using the chemical structure of chemical compounds as the sole input, one can build a toxicity prediction model based on parameters that define the physico-chemical properties and relative reactivity of the compounds. The objective of this thesis was to apply OECD guidelines in the development of validated QSAR models that describe acute toxicity of selected groups of EINECS chemicals to various organisms. In addition, an estimate was made of the total number of EINECS chemicals that could be possibly evaluated using (Q)SAR approaches.

Based on experimental toxicity data from literature and *in silico* calculated log K_{ow} (a measure of hydrophobicity) values, a QSAR advisory tool was developed that directs users to the appropriate QSAR model to apply for predicting toxicity of substituted mononitrobenzenes to five types of organisms within specified log K_{ow} ranges. In a next study, QSAR models were developed to predict *in vivo* acute toxicity of chlorinated alkanes to fish based on data from *in vitro* experiments, and even based on *in silico* log K_{ow} data only. Furthermore, using toxicity data from acute immobilization experiments with daphnids, an interspecies QSAR model was developed to predict toxicity of organothiophosphate pesticides to fish based on toxicity data for daphnids and *in silico* log K_{ow} values. The QSAR models for the mononitrobenzenes, chlorinated alkanes, and organothiophosphates covered in total 0.7 % of the 100 196 EINECS chemicals. In a final step, using chemical classification software, 54 % of the EINECS chemicals were grouped into specific classes that can in theory be subject to QSAR modeling. The safety assessment of one group of compounds that could not be classified e.g. botanical extracts might be done by further development of a method

recently reported for the safety assessment of natural flavour complexes used as ingredients in food. This would result in an additional 3 % of the EINECS chemicals that could potentially be covered by SAR approaches, bringing the total percentage of EINECS compounds that can be covered by (Q)SAR approaches to 57.

In conclusion, the results of this thesis reveal that, (i) *in vitro* experiments and even *in silico* calculations can help to reduce or replace animals used for experimental toxicity testing and (ii) despite the fact that individual QSARs may often each cover only limited, i.e. less than 1%, of the EINECS compounds, (Q)SAR approaches have the potential to cover about 57 % of the EINECS compounds.

Table of contents


1

General introduction, objectives, and outline of thesis

Introduction

Chemicals are part of everyday life. They are present in our shower gels, the shoes we wear, the medicines, textiles and cars we use, the packaging material of our food, the electronic devices that make our life convenient and many more. The importance of chemicals in everyday life is such that we cannot do without them and their use is ever increasing. Global trade in chemicals almost doubled from € 962 billion in 1996 to € 1 641 billion in 2006 (*1*). The chemical industry within the European Union (EU) accounts for about € 476 billion of the total sales (Figure 1.1), which is about 30 % of the global sales. The EU chemical industry comprises about 27 000 enterprises employing about 1.2 million people, which accounts for about 6 % of the total workforce in the manufacturing industries (*1*).

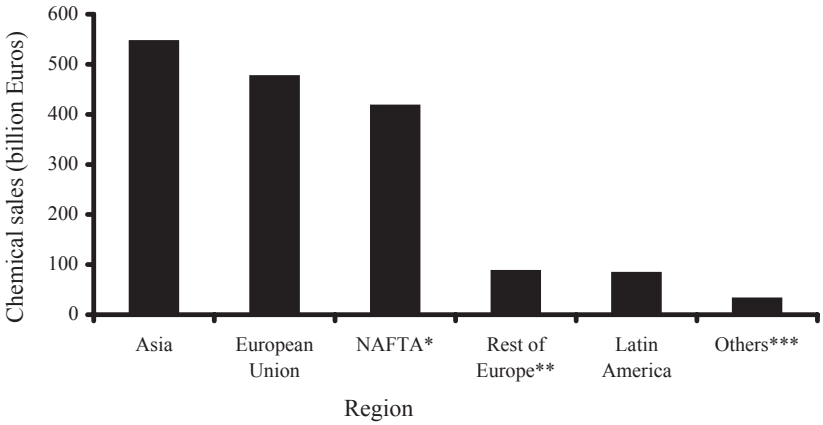


Figure 1.1: Geographical breakdown of world chemical sales. *North American Free Trade Area, **Switzerland, Norway and Central and Eastern Europe, ***Oceania and Africa. Adapted from CEFIC, 2007)

The use and production of chemicals has been associated with several industrial and environmental incidences over the years. These include for example, (i) the methylmercury poisoning in the Minamata bay in Japan between 1932 and 1968 (*2, 3*), (ii) the dichloro-diphenyl-trichloroethane (DDT) pesticide poisoning in the USA in the 1960s, documented in the book, Silent Spring, by Rachel Carson (*4*), (iii) the dioxin poisoning incident in Seveso, Italy in 1976 (*5*), and (iv) the methyl isocyanate poisoning in Bhopal, India, in 1984, popularly known as the Bhopal incident (*6, 7*). In addition, there is chronic exposure to much lower levels of chemicals,

which is much more often occurring than dramatic incidences. Incidences of some diseases e.g. testicular cancer in young men, breast cancer, and allergies have increased significantly over the last decades and some evaluations have related this to increased exposure to chemicals (*8*). The Scientific Committee on Toxicity, Ecotoxicity and the Environment (CSTEE) stated that "…links have been reported between high levels of persistent potential endocrine disrupting chemicals and reproductive and developmental effects in wildlife populations" (*9*).

Globally, there are about 16 million chemicals registered for use by the Chemical Abstracts Service. Within the EU, the registration of chemicals is governed by regulation (EC) 793/93, which classifies chemicals into two categories, "existing" or "new", depending on the date on which they entered the market (*10*). "Existing" chemicals were already on the EU market in the period between 1 January 1971 and 18 September 1981. They amount to 100 196 chemicals and are listed on the European Inventory of Existing Commercial Chemical Substances (EINECS) (EC, 1990). Since 19 September 1981 only 4 381 "new" chemicals (at the time of writing this thesis) have been added and they are listed on the European List of Notified Chemical Substances (ELINCS) (*11*). In comparison, in the USA, about 2 000 chemicals are notified every year. Both ELINCS and EINECS chemicals are regulated by more than 40 pieces of legislation (*12*). While ELINCS chemicals were required to be tested before introduction on the market when their volumes were as low as 10 kg per year and higher, EINECS chemicals were exempted from premarket testing, even though they constituted 99 % of the total volume of chemicals on the market at that time (*8*). Therefore, there exist large toxicity data gaps for EINECS chemicals. Out of the approximately 2 700 high production volume (HPV; production volume > 1 000 ton per year) EINECS chemicals, 141 chemicals were identified as priority substances for risk assessment and possible recommendations for risk reduction (*12*). Comprehensive risk evaluation has been completed for less than 30 % of these 141 chemicals. Analysis of the availability of data for all the HPV chemicals indicated that only 3 % of the HPV chemicals had a full toxicity data set needed for risk evaluation, while 28 % had minimal data sets (Table 1.1) (*13*).

To address the concerns about the lack of toxicity information for EINECS chemicals, the European Commission adopted a new regulatory framework for the Registration, Evaluation, and Authorization of Chemicals (REACH) in December 2006 (*14*). REACH came into force in June 2007 and it shifts the responsibility of assessing and managing the risks posed by chemicals from governments to industry.

Table 1.1: Percentage of the 2 700 high production volume EINECS chemicals for which effect data are available. (*13*)

Test type	Full[a] dataset (%)	Minimal[b] dataset (%)
Environmental Fate/Biodegradation	30	61
Ecotoxicological	9	72
Acute toxicity	29	81
Chronic toxicity	58	58
Mutagenicity	37	68
Developmental/Reproductive	20	15
Environment	5	57
Human Health	12	33
All data	3	28

[a]*Full dataset* –Environmental fate (photodegradation, stability in water and soil, biodegradation, bioaccumulation)
Ecotoxicological (toxicity to aquatic invertebrates, aquatic plants, andmicroorganisms)
Acute toxicity (oral, inhalation, dermal, eye and skin irritation, sensitization)
Chronic toxicity (repeated dose toxicity)
Mutagenicity (genetic toxicity *in vitro* and *in vivo*)
Developmental and reproductive toxicity
[b]*Minimal dataset* –Any one of the endpoints in environmental fate, ecotoxicity, acute toxicity, mutagenicity and developmental/reproductive toxicity plus biodegradation and chronic toxicity

Some of the main aims of REACH are (i) improved protection of human health and the environment, (ii) maintenance and enhancement of the competitiveness of the EU chemical industry, and (iii) promotion of non-animal alternatives to chemical testing. A comparison of the old and new (REACH) chemical management systems is shown in Table 1.2. Within the REACH framework, manufacturers are required by the year 2018 to provide toxicity information for about 30 000 chemicals with annual production volumes greater than 1 ton. The amount of toxicity data required for a chemical increases with increasing production volume.

There is currently ongoing debate about the potentially large number of animals that have to be used for experimental toxicity testing as an outcome of REACH. These concerns also arise from the numbers of animals that are currently used worldwide for experimental purposes. In 2005, about 14 million animals were reported to be used worldwide, with use within the EU accounting for 12 million (*15*), and the USA using about 2 million animals (*16*). The USA figure excludes rats and mice, because their usage is not generally reported (*16*). The numbers of animals used could be higher than 100 million per year, as usage is not reported on other continents. The guiding principles for animal experimentation over the last 50 years are mainly based on measures aimed at reducing, refining, and/or replacement of animals for experimental purposes, traditionally known as the "three R's" (*17*).

Table 1.2: Comparison of the old EU chemical management system, and the new system, REACH. (*14*)

Old chemical management system	REACH
Knowledge gaps in many of the chemicals on the EU market.	Close the knowledge gaps by providing safety information about chemicals produced or imported in volumes higher than 1 ton /year per manufacturer / importer.
The 'burden of proof' was on the authorities who needed to prove that the use of a chemical substance was unsafe before they could impose restrictions.	The 'burden of proof' will be on industry and all actors in the supply chain who have to demonstrate that the chemical can be used safely.
Notification requirements for 'new substances' started at a production level of 10 kg. Already at this level, one animal test was needed. At 1 ton, a series of tests including other animal tests had to be undertaken.	Registration will be required when production/ import reaches 1 ton. As far as possible, animal testing will be minimized.
Costly to introduce a new substance on the market. This encouraged the continued use of "existing", untested chemicals and inhibited innovation.	Innovation of safer substances will be encouraged under REACH through lower registration costs for new substances.
Public authorities were obliged to perform comprehensive risk assessments that were slow and cumbersome.	Industry will assess the safety of identified uses, prior to production and marketing. Authorities will focus on issues of serious concern.

These principles encourage (i) replacement of the use of animals with alternative techniques, (ii) reducing the number of animals used to a minimum, obtaining information from fewer animals or more information from the same number of animals, and (iii) refining the way experiments are carried out to ensure that animals suffer as little as possible (*17*).

There have been global efforts to implement the three R's while still better monitoring the effects of chemicals. Efforts to implement the three R's have resulted in the establishment of animal alternative centers, e.g. the John Hopkins Center for Alternatives to Animal Testing (CAAT; http://caat.jhsph.edu/) in USA, the Netherlands Centre for Alternatives (NCA; http://www.vet.uu.nl/nca/), and the European Center for the Validation of Alternative Methods (ECVAM; http://ecvam.jrc.it/). Within the REACH framework, there are provisions to allow alternative testing methods e.g. *in vitro* and *in silico*. Due to the large amount of toxicity information that is required within 10 years, methods that can help to direct priorities for future toxicity testing will help to save time, money, and animals. In this regard, *in silico* methods such as those based on (Quantitative) Structure Activity Relationships

[(Q)SARs] are expected to provide toxicity information or identify hazards of chemicals, and this information can subsequently be used to identify priority chemicals for further risk evaluation. The objective of this thesis was to develop validated QSAR models that can be used to predict the toxicity of chemicals, and thus help set priorities for future toxicity testing of chemicals and help to reduce the numbers of animals used for experimental purposes within REACH. In addition, an estimate was made of the total number of EINECS chemicals that could be possibly evaluated using QSARs.

(QUANTITATIVE) STRUCTURE-ACTIVITY RELATIONSHIPS

A (Quantitative) Structure-Activity Relationship [(Q)SAR] is based on the assumption that the biological activity of a new or untested chemical can be inferred from the molecular structure, or properties of similar compounds whose activities have already been assessed (*18*). When the relationship is developed with (i) non continuous or categorical data, it is called a SAR and (ii) continuous or quantitative data it is called a QSAR (*18*). In order for one to perform a QSAR analysis, three elements are needed: (i) biological data for a set of chemicals, (ii) a descriptors e.g. for physical or chemical properties of the chemicals, and (iii) a statistical method to relate the biological activity and the descriptor(s) (*19*). The two main fundamental assumptions of QSARs are that (i) similar chemicals have sufficiently common mechanistic elements, that they may have a rate-determining step and free energy requirements for activity in common, and (ii) differences in reaction rates for this common rate-limiting step will give rise to observed differences in activity or quantitative potency (*18*).

The first work on QSARs can be traced back to the PhD thesis work of Cros (1863) in Strasbourg, France, who showed that a relationship exists between the toxicity of primary aliphatic alcohols and their water solubility (*20*). In the field of aquatic toxicology, QSARs can be traced back more than 100 years to the work of Hans-Hörst Overton in Zürich and Meyer in Marburg who demonstrated that the potency of substances producing narcosis in tadpoles and small fish is in direct proportion to their partition coefficients measured between olive oil and water (*21, 22*). In 1969, Corwin Hansch, considered the founder of modern QSAR research, stated that the biological activity for a group of congeneric chemicals is related to their hydrophobicity, electronic and steric properties (*23*). In 1970, McFarland modified this relationship and indicated that toxicity is a combination of uptake into, or through, biological membranes (log of penetration) and the interaction of the toxicant with the molecular site of action (log of interaction) (Eq. 1), taking into account both the

toxicokinetics and toxicodynamics of the compound (*24*).

log toxicity = a(log of penetration) + b(log of interaction) Eq. 1

where a and b are reaction coefficients.

The process of developing a QSAR model involves several basic steps (Figure 1.2).

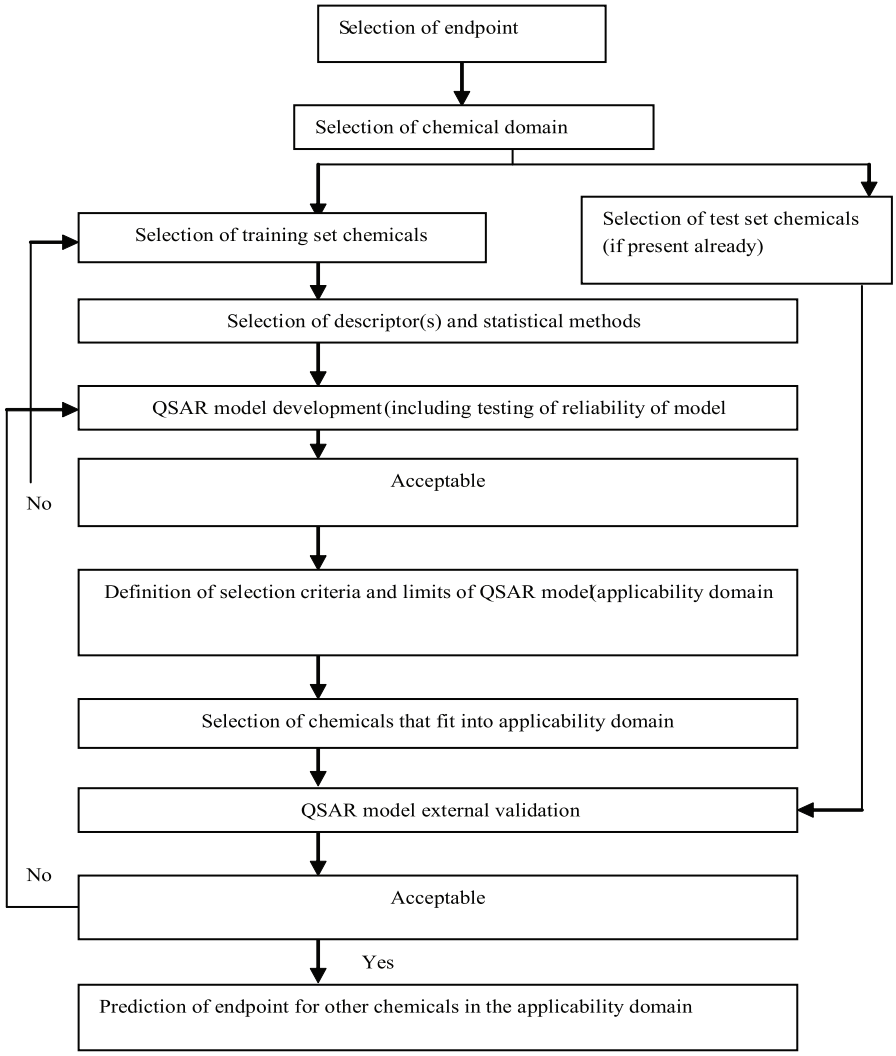


Figure 1.2: Basic scheme for the development of QSAR models. Adapted from (*19*). Modified to include OECD guidelines for QSAR model development (*25*).

Initially an endpoint to be modeled is selected, followed by identification of the chemical group for which the model will be developed. If the number of chemicals is sufficiently large, they can be split into a training set and a test set. Training set chemicals are used during the development of the model, while the test set chemicals are used to validate the model. The selection of appropriate descriptors to relate to the endpoint can be done *a priori* if the mechanism of action is already known, or one can start with a large number of descriptors and use computer software to select the best descriptors that correlate with toxicity. The descriptors can be physicochemical (e.g. hydrophobic), electronic (e.g. dipole moment, energy of the highest occupied molecular orbital (E_{homo}) or the energy of the lowest unoccupied molecular orbital (E_{lumo}) or steric (molecular volume, molecular weight) (*19*). Examples of commonly used descriptors and the xenobiotic characteristic they reflect are shown in Table 1.3.

Table 1.3: Examples of descriptors that can be calculated and the relevant xenobiotic characteristics.

Calculated descriptor	Relevant xenobiotic characteristic	Reference(s)
octanol water partition coefficient; log K_{ow}	hydrophobicity/lipophilicity	(*26, 27*)
dissociation constant; pKa	ionization	*(28)*
energy of the highest occupied molecular orbital; E_{homo}	ionization potential, ease of oxidation, nucleophilic reactivity	(*29, 30*)
energy of the lowest unoccupied molecular orbital; E_{lumo}	oxidation potential, ease of reduction, electrophilic reactivity	*(29)*
molecular weight, molecular volume, molecular surface area	binding to receptor	(*31, 32*)
molecular refractivity; MR	size and polarizability of a molecule fragment	*(33)*
heat of formation	activation or reaction enthalpies	*(29)*
dipole moment	charge separation in a molecule	*(29)*

The most commonly used physicochemical descriptor is the octanol-water partition coefficient, log K_{ow}, which reflects the ability of organic compounds to passively partition and accumulate in organisms (*26*). The importance of hydrophobicity in explaining the toxicity of a large set of industrial chemicals was shown for example, by the work of Könemann and Veith (*26, 27*). Where log K_{ow} alone is not sufficient to account for toxicity of chemicals, additional descriptors to be included may be (among others), electronic ones e.g. the energies of the frontier orbitals of the molecules (Figure 1.3).

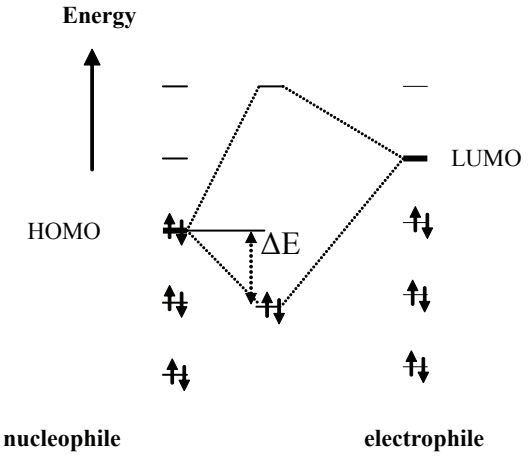


Figure 1.3: Molecular orbital diagram for the reaction between an electrophile and a nucleophile of a reactive toxic intermediate and its toxicological receptor. (*29*). The smaller the difference in energy between the HOMO of the nucleophile and the LUMO of the electrophile, the larger the energy ΔE gained as a result of the interaction and the higher the reactivity (*34*). ΔE = energy gain from the interaction between HOMO of the nucleophile and LUMO of the electrophile

The energies of the frontier orbitals e.g. the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital (HOMO) direct the electrophilic and nucleophilic reactivity of a compound and its toxicological receptor (*34*). Reactivity between an electrophile and a nucleophile increases when i) the E_{homo} is increased or ii) the E_{lumo} is decreased (*34*). Given that the toxicological receptor is constant for a series of chemicals to be modeled by a QSAR, the relative reactivity and thus toxicity of a series of chemicals may be modeled by looking at their relevant frontier orbital without the requirement for knowledge on the orbital characteristics of the toxicological receptor.

Traditional QSARs use experimentally derived descriptors e.g. ionization potential, vapor pressure, and log $K_{ow,}$ among others to quantify physico-chemical characteristics (*35, 36*). However, due to lack of large data sets of experimentally derived parameters, QSARs have been developed based on parameters defined using quantum mechanical computer calculations. Due to the increasing power of computers, quantum mechanical computer calculations have become a valuable and widely applicable tool in (bio)chemical research. Using the chemical structure of a compound as the sole input, they provide the possibility to calculate parameters that define the physico-chemical properties and relative reactivity of a compound. The possibilities for the use of computational quantum mechanical

calculations in QSAR studies in toxicology have already been demonstrated. (*29, 37-43*). Computational models offer the advantages of ease of use, speed, and low costs.

The correlation between the chosen descriptor(s) and the endpoint is often analyzed with statistical software. The most common correlative method is regression analysis. Regression analyses are simple to carry out and produce results that are easy to understand (*18, 44*). The correlation can also be investigated using pattern recognition techniques, and they are usually multi-dimensional and non-linear, thus they are often complex and difficult to interpret (*18*).

In a next step, the reliability of the developed QSAR model can be tested by comparing the endpoint values the model predicts to the experimentally determined endpoint values of similar chemicals (training set chemicals). If the predictions are poor, one can restart the model development by using different descriptors, or noting training set compounds, whose predicted values deviate greatly from the experimental values. This can help to identify compounds that act by a different mechanism of action (*19*). If the predictions are good, one can define the selection criteria and limits of the model and then make toxicity predictions to other chemicals that meet the selection criteria, and were not used during model development.

Although QSAR models have been used in regulatory assessment of chemical safety in many countries for many years, there were no universal principles for their regulatory applicability. Researchers used different criteria for judging the quality of QSAR models. In 2004, member countries of the Organization for Economic Cooperation and Development (OECD) agreed on the principles for developing and validating (Q)SAR models for their use in regulatory assessment of chemical safety (*45*). These guidelines are discussed in brief in the next section.

OECD GUIDELINES FOR (Q)SAR DEVELOPMENT AND VALIDATION

In order to facilitate the consideration of a (Q)SAR model for regulatory purposes, the (Q)SAR model should be associated with the following five categories of information: i) a defined endpoint, ii) an unambiguous algorithm, iii) a defined domain of applicability, iv) appropriate measures of goodness-of-fit, robustness, and predictivity, and v) a mechanistic interpretation (*45*). In 2007, the OECD published a "Guidance Document on the Validation of (Q)SAR Models" with the aim of providing guidance on how specific (Q)SAR models can be evaluated with respect to the OECD principles. In the following section, each of the five categories is discussed briefly.

i) A defined endpoint

A QSAR model needs to be based on a clearly defined end-point that refers to a physicochemical, biological, or environmental effect that can be experimentally determined and thus modeled. This implies that it should be not only clear which endpoint is being modeled but also how it is experimentally defined, since a given endpoint could be determined under different experimental conditions and protocols (*25*).

ii) An unambiguous algorithm

An unambiguous algorithm seeks to ensure transparency of the model that is used to generate predictions of a toxicity endpoint based on chemical structure and/or physicochemical properties (*25*). This holds for the algorithm defining the QSAR model but also for the descriptors used in these algorithms. Descriptors relevant to the toxicity endpoint, calculation methods, software packages, and algorithms should preferably be based on methods that are publicly available. The statistical technique, used to analyze the relationship between toxicity and the physicochemical descriptors should be transparent (*25*).

iii) A defined domain of applicability

As models are a simplification of reality, their limits should be well defined (*46*). This principle reflects the fact that QSARs are associated with limitations regarding the types of chemical structures, physicochemical properties, and mechanism for which a model generates reliable predictions. In the simplest approach the domain of applicability can be defined by the boundaries of the descriptor values for the compounds in the training set (*47*).

iv) Appropriate measures of goodness-of-fit, robustness, and predictivity,

This principle requires that parameters that reflect both the internal performance of the QSAR model as well as its predictivity should be provided. The internal performance can be characterized based on the goodness-of-fit and robustness of the model determined based on the training set. These qualities of the QSAR models can be characterized by the number of compounds used in the study (n), coefficient of determination (r^2), standard error of the estimate (s), variance ratio (F), the internally cross validated coefficient of determination (r^2_{int}), and the externally validated co-efficient of determination (r^2_{ext}) (*25*).

v) A mechanistic interpretation

The OECD guidelines also state that a QSAR should be associated with a mechanistic interpretation, whenever this is possible. Such a mechanistic interpretation links the descriptors

used in the model and the endpoint being predicted (*25*). Any effort in the validation process to show that the QSAR model is consistent with other knowledge of fundamental processes in chemistry and toxicology increases the credibility and acceptance of the predictions from the model (*25*). The mechanism(s) of toxicity of chemicals are broadly classified into four categories: i) non-polar narcotics, ii) polar narcotics, iii) unspecific reactivity, and (iv) specific mechanism of action (*48*). Non-polar narcotics do not interact with any specific receptors in an organism and their toxicity is entirely dependent on their hydrophobicity. In the absence of any specific mechanism of toxicity, a chemical will always be as toxic as its hydrophobicity indicates, a phenomenon called "baseline toxicity". Polar narcotics are slightly more toxic than baseline toxicity, and they usually possess hydrogen bond donor activity, e.g. phenols and anilines (*49*). Chemicals with unspecific or specific reactivity have enhanced toxicity when compared to baseline toxicity. Unspecifically reacting chemicals react with certain chemical structures commonly found in biomolecules, e.g. epoxides, which react with sulfhydryl groups of cysteine residues of peptides while specifically acting chemicals react with specific receptor molecules e.g. DDT, which interacts with sodium channel regulating receptors in neurons (*48*).

Altogether, these OECD guidelines should provide regulatory bodies with a scientific basis for making decisions on the acceptability of data generated by (Q)SARs, and also promote the mutual acceptance of (Q)SAR models by improving the transparency and consistency of (Q)SAR reporting.

Outline of thesis

The aim of this thesis was to develop a computational chemistry-based QSAR approach that enables identification of priorities within various selected groups of EINECS chemicals. Validated QSAR models for acute toxicity of selected groups of EINECS chemicals were developed taking into account the OECD guidelines (*25*). **Chapter 1** gives a general introduction on the subjects that are relevant within the context of the present thesis. In the next three chapters, QSAR models were developed for mononitrobenzenes **(chapter 2)**, chlorinated alkanes **(chapter 3)** and organothiophosphate pesticides **(chapter 4)**. Suitable software packages, descriptor calculation protocols, and statistical techniques for use for the rest of the thesis were also identified in **chapter 2**. In each of the three chapters **(2, 3,** and **4)**, the number and type of EINECS chemicals for which the developed QSAR models were valid were identified. In a final chapter an estimate was made of the percentage of EINECS chemicals that can be grouped into specific chemical classes, and thus in theory be subject

to QSAR modeling **(chapter 5)**. This gives an indication of the potential applicability of QSAR models to predict acute toxicity of chemicals within REACH. Finally, the overall conclusions and a general discussion of this thesis are presented **(chapter 6)**.

References

(1) CEFIC. (2007) Facts and figures: The European Chemical industry in a worldwide perspective, available online at http://www.cefic.be/factsandfigures/.

(2) Okahata, T. (1967) Minamata disease and its symptoms. *[Hokenfu zasshi] The Japanese journal for public health nurse 23*, 43-47.

(3) Takeuchi, T., Morikawa, N., Matsumoto, H. and Shiraishi, Y. (1962) A pathological study of Minamata disease in Japan. *Acta Neuropathologica 2*, 40-57.

(4) Carson, R. (1962) *Silent Spring*, Fawcett Crest, New York.

(5) Homberger, E., Reggiani, G., Sambeth, J. and Wipf, H. K. (1979) The Seveso accident: Its nature, extent and consequences. *Annals of Occupational Hygiene 22*, 327-367.

(6) Mehta, P. S., Mehta, A. S., Mehta, S. J. and Makhijani, A. B. (1990) Bhopal tragedy's health effects. A review of methyl isocyanate toxicity. *Journal of the American Medical Association 264*, 2781-2787.

(7) Bucher, J. R. (1987) Methyl isocyanate: a review of health effects research since Bhopal. *Fundamental and Applied Toxicology 9*, 367-379.

(8) EU. (2001) White Paper: Strategy for a future chemicals policy. Commission of the European Committees, Brussels, Belgium. COM(2001) 88 final., pp 1–32.

(9) CSTEE. (1999) CSTEE Opinion on Human and Wildlife Health Effects of Endocrine Disrupting Chemicals, with Emphasis on Wildlife and on Ecotoxicology Test Methods, In *Report of the Working Group on Endocrine Disrupters of the Scientific Committee on Toxicity, Ecotoxicity and the Environment (CSTEE) of DG XXIV, Consumer Policy and Consumer Health Protection*.

(10) EU. (1993) Council Regulation (EEC) No 793/93 of 23 March 1993 on the evaluation and control of the risks of existing substances *Off. J. Eur. Union. L 084*, 1–75.

(11) ELINCS. European List of Notified Chemical Substances; available online at http://ecb.jrc.it/esis/index.php?PGM=eli).

(12) EU. (2007) REACH in brief, European Commission; Enterprise and Industry Directorate General, Brussels.

(13) Allanou, R., Hansen, B. G. and van der Bilt, Y. (1999) Public availability of data on EU high production volume chemicals, pp 1–21, European Commission, JRC, Institute for Health and Consumer Protection, European Chemicals Bureau, Italy.

(14) EU. (2006) Regulation (EC) No 1907/2006 of the European Parliament and of the Council. *Off. J. Eur. Union*, 1–849.

(15) EU. (2007) Fifth report on the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union, pp 1-14, Brussels.

(16) Kulpa-Eddy, J., Snyder, M. and Stokes, W. (2008) A review of trends in animal use in the United States (1972-2006), In *6th World Congress on Alternatives and Animal Use in the life sciences* pp 163-165, AATEX, Tokyo, Japan.

(17) Russell, W. M. S. and Birch, R. (1959) *The principles of humane experimental technique*, Methuen, London.

(18) Schultz, T. W., Cronin, M. T. D., Walker, J. D. and Aptula, A. O. (2003) Quantitative structure-activity relationships (QSARs) in toxicology: a historical perspective. *Journal of Molecular Structure: THEOCHEM 622*, 1-22.

(19) Walker, J. D., Jaworska, J., Comber, M. H. I., Schultz, T. W. and Dearden, J. C. (2003) Guidelines for developing and using Quantitative Structure Activity Relationships. *Environmental Toxicology and Chemistry 22*, 1653-1665.

(20) Cros, A. F. A. (1863) Action de l'alcohol amylique sur l'organisme, In *Faculty of Medicine*,

University of Strasbourg, Strasbourg.

(21) Overton, E. (1901) *Studien uber die Narkose, zugleich ein Beitrag zur allgemeinen Pharmakologie*, 1-195.

(22) Meyer, H. (1899) Zur Theorie der Alkoholnarkose - Erste Mittheilung. Welche Eigenschaft der Anasthetica bedingt ihre narkotische Wirkung. *Archiv fur Experimentelle Pathologie und Pharmakologie 42*, 109-118.

(23) Hansch, C. (1969) A quantitative approach to biochemical structure-activity relationships. *Accounts of Chemical Research 2*, 232-239.

(24) McFarland, J. W. (1970) On the parabolic relationship between drug potency and hydrophobicity. *Journal Of Medicinal Chemistry 13*, 1192-1196.

(25) OECD. (2007) Guidance document on the validation of (Quantitative) structure-activity relationships [(Q)SAR] models, ENV/JM/MONO(2007)2, In *OECD Environment Health and Safety Publications: Series on Testing and Assessment No. 69. Paris, France* pp 1–154.

(26) KÖnemann, H. (1981) Quantitative structure-activity relationships in fish toxicity studies Part 1: Relationship for 50 industrial pollutants. *Toxicology 19*, 209–221.

(27) Veith, G. D., Call, D. J. and Brooke, L. T. (1983) Structure - toxicity relationships for the fathead minnow, *Pimephales promelas*: narcotic industrial chemicals. *Can. J. Fish. Aquat. Sci. 40*, 743–748.

(28) Schultz, T. W., Lin, D. T. and Wesley, S. K. (1992) QSARs for monosubstituted phenols and the polar narcosis mechanism of toxicity. *Quality assurance (San Diego, Calif.) 1*, 132-143.

(29) Soffers, A. E. M. F., Boersma, M. G., Vaes, W. H. J., Vervoort, J., Tyrakowska, B., Hermens, J. L. M. and Rietjens, I. M. C. M. (2001) Computer-modeling-based SARs for analyzing experimental data on biotransformation and toxicity. *Toxicol in Vitro 15*, 539-551.

(30) Cnubben, N. H., Peelen, S., Borst, J. W., Vervoort, J., Veeger, C. and Rietjens, I. M. (1994) Molecular orbital-based quantitative structure-activity relationship for the cytochrome P450-catalyzed 4-hydroxylation of halogenated anilines. *Chem. Res. Toxicol. 7*, 590-598.

(31) Hou, T. J. and Xu, X. J. (2003) ADME Evaluation in Drug Discovery. 3. Modeling Blood-Brain Barrier Partitioning Using Simple Molecular Descriptors. *Journal of Chemical Information and Computer Sciences 43*, 2137-2152.

(32) Lesigiarska, I., Pajeva, I. and Yanev, S. (2002) Quantitative structure-activity relationship (QSAR) and three-dimensional QSAR analysis of a series of xanthates as inhibitors and inactivators of cytochrome P450 2B1. *Xenobiotica 32*, 1063-1077.

(33) Hansch, C., Jazirehi, A., Mekapati, S. B., Garg, R. and Bonavida, B. (2003) QSAR of apoptosis induction in various cancer cells. *Bioorganic and Medicinal Chemistry 11*, 3015-3019.

(34) Fleming, I. (1976) Molecular Orbitals and Frontier Orbitals., In *Frontier Orbitals and Organic Chemical Reactions*, John Wiley & Sons, New York.

(35) Abraham, M. H., Chadha, H. S., Martins, F., Mitchell, R. C., Bradbury, M. W. and Gratton, J. A. (1999) Hydrogen bonding part 46: A review of the correlation and prediction of transport properties by an LFER method: Physicochemical properties, brain penetration and skin permeability. *Pesticide Science 55*, 78.

(36) Hansch, C. and Fujita, T. (1964) A method for the correlation of biological activity and chemical structure. *J. Amer. Chem. Soc. 86*, 1616 - 1626.

(37) Barratt, M. D. (1998) Integrating computer prediction systems with in vitro methods towards a better understanding of toxicology. *Toxicol Lett 102-103*, 617-621.

(38) Cronin, M. T., Gregory, B. W. and Schultz, T. W. (1998) Quantitative structure-activity analyses of nitrobenzene toxicity to *Tetrahymena pyriformis*. *Chem. Res. Toxicol. 11*, 902-

908.

(39) Damborsky, J. and Wayne Schultz, T. (1997) Comparison of the QSAR models for toxicity and biodegradability of anilines and phenols. *Chemosphere 34*, 429-446.

(40) Friedig, A. P. and Hermens, J. L. (2000) Narcosis and chemical reactivity QSARs for acute fish toxicity. *Quant. Struct. Act. Relat. 19*, 547 - 553.

(41) Hatch, F. T. and Colvin, M. E. (1997) Quantitative structure-activity (QSAR) relationships of mutagenic aromatic and heterocyclic amines. *Mutat Res 376*, 87-96.

(42) Schmitt, H., Altenburger, R., Jastorff, B. and Schuurmann, G. (2000) Quantitative structure-activity analysis of the algae toxicity of nitroaromatic compounds. *Chem Res Toxicol 13*, 441-450.

(43) Worrall, F. and Thomsen, M. (2004) Quantum vs. topological descriptors in the development of molecular models of groundwater pollution by pesticides. *Chemosphere 54*, 585-596.

(44) Twomey, P. J. and Kroll, M. H. (2008) How to use linear regression and correlation in quantitative method comparison studies. *International Journal of Clinical Practice 62*, 529-538.

(45) OECD. (2004) The report from the expert group on (Quantitative) Structure-Activity Relationship ((Q)SARs) on the principles for the validation of (Q)SARS. ENV/JM/MONO(2004)24. , In *OECD Series on testing and assessment: number 49*.

(46) Worth, A. P. and Cronin, M. T. D. (2004) Report of the workshop on the validation of QSARs and other computational prediction models. *ATLA Alternatives to Laboratory Animals 32*, 703-706.

(47) Schultz, T. W., Cronin, M. T. D. and Netzeva, T. I. (2003) The present status of QSAR in toxicology. *Journal of Molecular Structure: THEOCHEM 622*, 23-38.

(48) Verhaar, H. J. M., van Leeuwen, C. J. and Hermens, J. L. M. (1992) Classifying environmental pollutants. *Chemosphere 25*, 471-491.

(49) Veith, G. D. and Broderius, S. J. (1990) Rules for distinguishing toxicants that cause type I and type II narcosis syndromes. *Environ. Health Persp. 87*, 207–211.

Quantum chemistry-based quantitative structure-activity relationships for modeling the (sub) acute toxicity of substituted mononitrobenzenes in aquatic systems

Elton Zvinavashe, Albertinka, J. Murk, Jacques Vervoort,
Ans E.M.F. Soffers, Andreas Freidig, Ivonne M.C.M. Rietjens

Environmental Toxicology and Chemistry (2006) 25: 2313–2321

Abstract

Fifteen experimental literature data sets on the acute toxicity of substituted nitrobenzenes to algae (*Scenedesmus obliquus*, *Chlorella pyrenoidosa*, *C. vulgaris*), daphnids (*Daphnia magna*, *D. carinata*), fish (*Cyprinus carpio*, *Poecilia reticulata*), protozoa (*Tetrahymena pyriformis*), bacteria (*Phosphobacterium phosphoreum*), and yeast (*Saccharomyces cerevisiae*) were used to establish quantum chemistry-based quantitative structure–activity relationships (QSARs). The logarithm of the octanol/water partition coefficient, log K_{ow}, and the energy of the lowest unoccupied molecular orbital, E_{lumo}, were used as descriptors. Suitable QSAR models ($0.65 < r^2 < 0.98$) to predict acute toxicity of substituted mononitrobenzenes to protozoa, fish, daphnids, yeast, and algae have been derived. The log K_{ow} was a sufficient descriptor for all cases, with the additional E_{lumo} descriptor being required only for algae. The QSARs were found to be valid for neutral substituted mononitrobenzenes with no -OH, -COOH, or -CN substituents attached directly to the ring. From the 100 196 European Inventory of Existing Commercial Substances (EINECS), 497 chemicals were identified that fit the selection criteria for the established QSARs. Based on these results, an advisory tool has been developed that directs users to the appropriate QSAR model to apply for various types of organisms within specified log K_{ow} ranges. Using this tool, it is possible to obtain a good indication of the toxicity of a large set of EINECS chemicals and newly developed substituted mononitrobenzenes to five different organisms without the need for additional experimental testing.

Introduction

The European Inventory of Existing Commercial Substances (EINECS) contains 100 196 chemicals that were on the European Union market between 1971 and September 1981. Any chemical marketed after September 1981 is called a ''new chemical'' (*1*). It is recognized that insufficient (eco)toxicological information exists on the hazardous properties of many of the EINECS chemicals. This is mainly because EINECS chemicals were exempted from premarket testing, unlike ''new chemicals,'' even though EINECS chemicals constituted 99% of the total volume of all substances on the market at that time (*1*). In October 2003, the European Commission adopted a proposal for a new chemical control system called Registration, Evaluation, and Authorization of Chemicals (REACH). One of the goals of REACH is the development of computational prediction models to fill existing toxicity data gaps (*1*). General information for about 30 000 existing chemicals marketed in the European Union in volumes greater than one ton per year is required before 2012 (*2*).

In the field of toxicological risk assessment, the present view is that validated quantitative structure–activity relationships (QSARs) can be valuable tools to rapidly screen the toxicological potential of chemical compounds and also to help set up research priorities in toxicological testing programs (*3, 4*). One of the principal assumptions that underlie the description of QSARs is that physical-chemical properties dominate the behavior of chemical compounds. Traditional QSARs use experimentally derived descriptors, such as ionization potential, vapor pressure, octanol/water partition coefficient (K_{ow}), and Hammett, Taft, Sterimol, and Abraham parameters to quantify these physical-chemical characteristics (*5, 6*). However, because of a lack of large data sets of experimentally derived parameters, attempts have been made to define parameters to establish QSARs based on quantum mechanical computer calculations. As the power of computers continues to increase, quantum mechanical computer calculations have become a valuable and widely applicable tool in (bio)chemical research. Quantum mechanical calculations provide the possibility to calculate, using the chemical structure of a compound as the sole input, parameters that define the physical-chemical properties and relative reactivity of a compound. The possibilities for the uses of computational quantum mechanical calculations in QSAR studies in toxicology have already been demonstrated. From these previous initial studies, several classes of compounds have been identified for which a possibility exists to use computational chemistry–based parameters to estimate their toxicity (*7-9*). Computational models offer the advantages of ease of use, speed, and low costs. Furthermore, they are in principle not susceptible to experimental errors, although slightly different output descriptors can sometimes result

because of numerical deviations or the different prediction algorithms used by the available software. In order to establish the accuracy of quantum chemistry-based data, they need to be rigorously validated with experimental data of good quality.

Currently, attempts are being made to define QSARs for several compound classes. Substituted nitrobenzenes are a good starting point for two reasons. First, sufficient experimental toxicity data exist, and, second, no attempt has been made to aggregate the data and use them to define priorities for future toxicity testing. Substituted nitrobenzenes are widely used in industry during the synthesis of dyes, explosives (e.g., trinitrotoluene), solvents, plastics, anilines, and various bioactive products, such as insecticides, pesticides, and pharmaceuticals (*10, 11*). Substituted nitrobenzenes are also found as by-products of fuel combustion in vehicles and power plants and as secondary pollutants from reactions with hydroxyl and nitrate radicals. As a result of their varied origins and uses, substituted nitrobenzenes are widespread in ecosystems and consequently have a high potential for causing ecotoxic effects (*10*). Substituted nitrobenzenes are generally electrophilic toxicants because of the presence of the strong electron-withdrawing nitro substituent (*9*). The presence of other substituents on the nitrobenzene ring results in different toxicities in part because of different susceptibilities to nitro-reduction by enzymes, giving rise to nitro radical anions, nitrosobenzenes, or *N*-hydroxylamines (*12*).

In this chapter, to estimate the sub(acute) toxicity of various substituted nitrobenzenes to aquatic organisms, QSARs were developed based on computer calculated descriptors, namely, the logarithm of the octanol/water partition coefficient (log K_{ow}) and the energy of the lowest unoccupied molecular orbital E_{lumo} (eV). The log K_{ow} generally models a compound's hydrophobicity, which is important in describing the passage of a compound through membranes. The E_{lumo} models the electrophilic nature of the nitrobenzenes, which is important in their nitroreduction and/or covalent reaction with biological macromolecules. The log K_{ow} and E_{lumo} have a clear link to the toxicity of substituted nitrobenzenes as shown from previous studies (*9, 13-15*). The applicability and limits of the QSARs were also identified by noting certain types of compounds that are outliers and certain log K_{ow} ranges where one should apply the QSARs for the non-polar narcotic models instead. Thus, from this, it is possible to estimate the acute toxicity of substituted mononitrobenzenes to several aquatic organisms. To this end, an advisory tool was established to direct the user to the appropriate QSAR for a given organism and log K_{ow} range.

MATERIALS AND METHODS

Data sets

The 15 experimental data sets used in this chapter contained acute toxicity data for substituted nitrobenzenes to six types of organisms. These were the protozoan *Tetrahymena pyriformis* (*9, 11*); algae *Scenedesmus obliquus* (*13, 16*), *Chlorella pyrenoidosa* (*15*), and *C. vulgaris* (*17*); daphnids *Daphnia magna* (*15*) and *D. carinata* (*13*); fish *Cyprinus carpio* (*13*) and *Poecilia reticulata* (*14*); bacteria *Phosphobacterium phosphoreum* (*15*); yeast *Saccharomyces cerevisiae* (*10*); and sub-acute toxicity data to *D. magna* (*15*). The European Chemicals Bureau supplied the EINECS list, from which the list of substituted nitrobenzenes falling into the applicability domain of the QSARs developed in this chapter was extracted.

Molecular descriptors

Seven commercially available software packages were compared for their accuracy in estimating log K_{ow} values for 77 substituted nitrobenzenes. The software packages used were CLogP Version 4.0 (Biobyte, Claremont, CA, USA) (*18*), ALogPs Version 2.102 (VCCLab, Munich, Germany) (*19*), K_{ow} WIN Version 1.66 (Syracuse Research, Syracuse, NY, USA) (*20*), IA LogP (Chem Silico, MA, USA; http://www.logp.com), Spartan 04 for Windows_ Version 1.0.3 (Wavefun, Irvine, CA, USA; http://www.wavefun.com),
SPARC online calculator (http://ibmlc2.chem.uga.edu/sparc/index.cfm), and ACD/LogP Version 4.67 (Advanced Chemistry Development, Toronto, ON, Canada) (*21*). The estimated log K_{ow} values were compared to experimental values obtained from the PHYSPROP database (Syracuse Research; http://www.syrres.com/esc/physdemo.htm). To input the structure of each molecule into the software, the simplified molecular input line entry system (SMILES) code was used. The SMILES codes are widely used to represent a molecular structure by a linear string of symbols (*22*). The SMILES codes were obtained from the SMILES-CAS database (Syracuse Research). The E_{lumo} values were calculated using a semi-empirical method, applying the Austin Model 1 (AM1) Hamiltonian (*23*) from the program Spartan 04 for Windows® Version 1.0.3 (Wavefun) for Windows®, v1.0.3, on a Pentium 4 computer with Windows XP Professional Service Pack 2 and 1 GB memory.

Statistical analyses

Linear regression analysis, performed in Microsoft® Excel 2000 (Microsoft, Redmond, WA, USA), was used to select the most suitable log K_{ow} estimation software for the substituted nitrobenzenes. Experimental log K_{ow} was plotted against log K_{ow} predicted from the different

software packages, and the resulting coefficient of determination (r^2) and standard error of the estimate (SE) values were recorded. The Statistical Package for Social Scientists (SPSS®) 10.1 for Windows® (SPSS, Chicago, IL, USA) was used to analyze the QSARs. For QSAR descriptions, log (toxicity endpoint) was the dependent variable, while log K_{ow} and E_{lumo} were the independent variables. All QSAR equations were presented in the form

$$\log \text{toxicity } (\mu M) = x(\log K_{ow}) + y(E_{lumo}) + z \qquad \text{(Eq. 1)}$$

where *x, y,* and *z* are fitting parameters. The quality of the QSAR models was characterized by the number of compounds used in the study (*n*), r^2, *s*, variance ratio (*F)*, and the internally cross-validated coefficient of determination (r^2_{int}). Cross validation of the models was done using a leave-out-many method, with 20% of the calibration compounds left out at each step. Cross validation was done only for models with $r^2 > 0.65$ and $n/k \geq 5$, where *k* is the number of descriptors (*24*). The models were considered acceptable for further use when $r^2_{int} \geq 0.5$ and $r^2 - r^2_{int} < 0.3$ (*25*). The validation groups were created using the method of unsupervised stratification of cross validation to reduce bias introduced by random sampling (*26*). The data were ranked according to increasing log K_{ow} values, then the compounds were alternately classified into five groups, and r^2_{int} was calculated according to the formula

$$r^2_{int} = 1 - (\text{PRESS/SSD}) \qquad \text{(Eq. 2)}$$

where predictive sum of squares (PRESS) is the sum of the squared differences between actual and predicted toxicity when the compounds are omitted from the regression and SSD is the sum-of-squares deviation for each actual toxicity from the mean toxicity of all the compounds (*27*). The chemical applicability domain of the models was defined in three ways. First, the degree of extrapolation was defined by the average leverage value, *h,* where leverage is a measure of the distance of a compound from the model experimental space. The warning leverage, h_w, was set at three times the average leverage value. Any compound with $h > h_w$ falls outside the optimum prediction space of the model (*24*). Second, external predictions are restricted to substituted nitrobenzenes that contain only the substituents present in the calibration set. Third, log K_{ow} cutoff points were set by comparing our QSAR equations to those for non-polar narcotics. Above the cutoff point, the non-polar narcotics QSAR should be applied in order to avoid underestimating the toxicity of compounds within the chemical domain of the QSAR (*28*).

RESULTS

Theoretically determined descriptors

The 15 data sets used to obtain experimental toxicity data to establish QSARs contained 103 substituted nitrobenzenes. The E_{lumo}, log K_{ow} values (estimated and experimental) for these substituted nitrobenzenes are presented as an appendix in Table S1 available online *(42)*. The E_{lumo} values ranged from -2.67 eV to -0.79 eV, while the estimated log K_{ow} values ranged from -0.14 to 5.03. Because of the large size of the experimental toxicity data sets, they are also presented as appendices in Tables S2 to S10 available online *(42)*.

Log K_{ow} estimation

Out of the 103 substituted nitrobenzenes, experimental log K_{ow} values were available for 77 (Table S1). The following values *($r^2 \pm s$)* were obtained for the various software packages: Spartan (0.83 $\pm$ 0.38), SPARC (0.87 $\pm$ 0.33), IA LogP (0.88 $\pm$ 0.32), ACD/LogP (0.94 $\pm$ 0.15), ALogPs (0.95 $\pm$ 0.21), K_{ow}WIN (0.96 $\pm$ 0.18), and CLogP (0.98 $\pm$ 0.15). A comparison of the highest r^2 value (CLogP) with the other r^2 values resulted in a statistically significant difference ($p < 0.05$) only with the r^2 values for SPARC and Spartan. Based on the highest r^2 value, the log K_{ow} values estimated using CLogP were used to establish QSARs.

Definition of selection criteria for QSARs

The acute toxicity data set of substituted nitrobenzenes to the ciliated protozoa *T. pyriformis* (*11*) was considered a suitable starting point for QSAR development, as it contains substituted nitrobenzenes with a wide variety of substituents ranging from halogen, carboxyl, hydroxyl, oxime, cyano, amide, to phenyl, all in the *ortho, para,* and/or *meta* positions. Figure 2.1a presents a plot of the decadic logarithm of the experimental concentration that inhibits growth by 50 % (IGC_{50}) for these substituted nitrobenzenes against values predicted based on the estimated log K_{ow}.

From Figure 2.1a, it can be seen that the following compounds are outliers: dinitrobenzenes, nitrophenols, nitrobenzoic acids, nitrodiphenyls, and cyanonitrobenzenes. The absolute standardized residuals for these compounds were all above the set limit of 1.5, confirming their outlier status. Selection criteria were established for the QSARs, therefore, which exclude these compound classes. Note that nitrobenzoic acids and nitrophenols do not apply to the QSAR also because they are likely to be ionized at physiological pH (*28*). Repeating the QSAR according to the selection criteria resulted in an improved correlation ($r^2 = 0.82$) when compared to the QSAR that included all the nitrobenzenes ($r^2 = 0.27$).

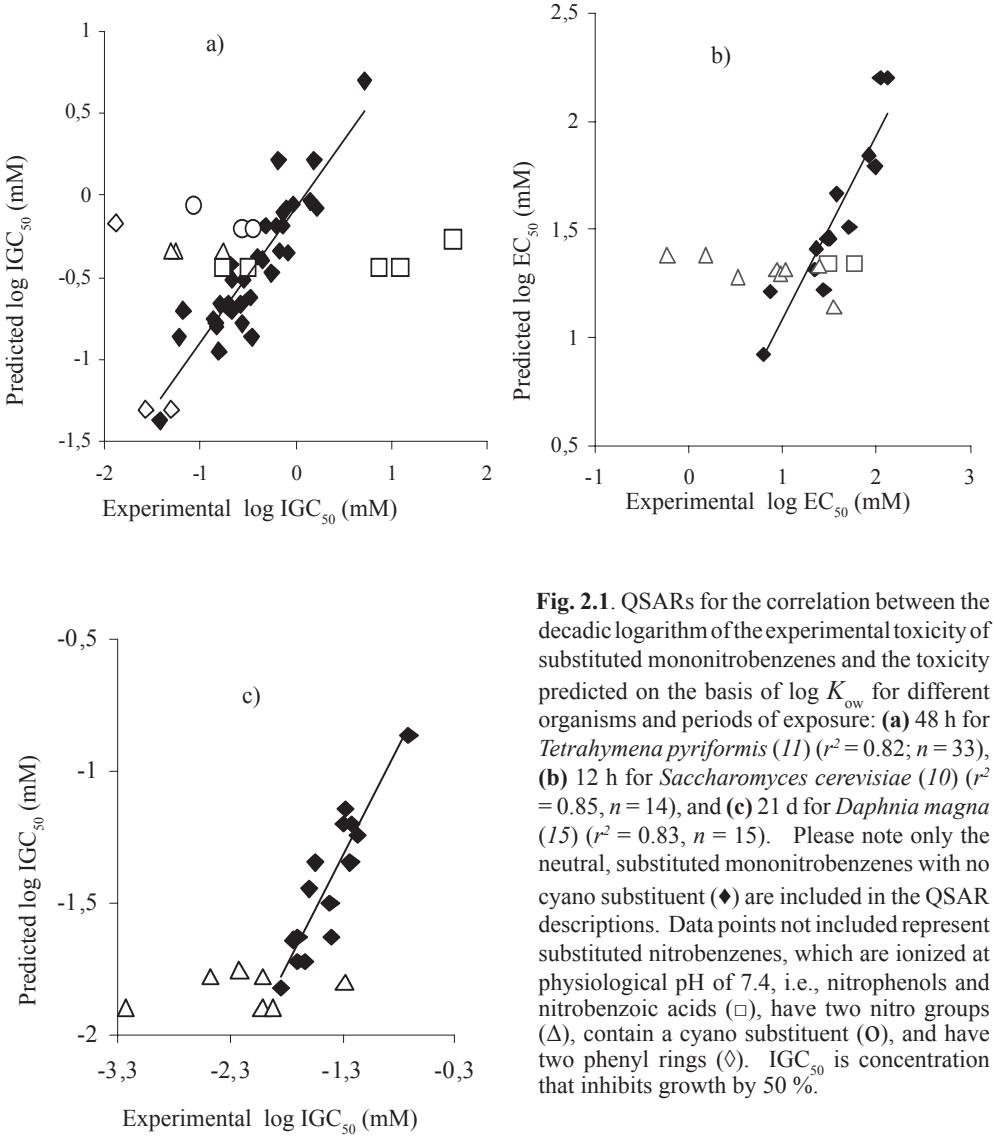


Fig. 2.1. QSARs for the correlation between the decadic logarithm of the experimental toxicity of substituted mononitrobenzenes and the toxicity predicted on the basis of log K_{ow} for different organisms and periods of exposure: **(a)** 48 h for *Tetrahymena pyriformis* (*11*) (r^2 = 0.82; n = 33), **(b)** 12 h for *Saccharomyces cerevisiae* (*10*) (r^2 = 0.85, n = 14), and **(c)** 21 d for *Daphnia magna* (*15*) (r^2 = 0.83, n = 15). Please note only the neutral, substituted mononitrobenzenes with no cyano substituent (♦) are included in the QSAR descriptions. Data points not included represent substituted nitrobenzenes, which are ionized at physiological pH of 7.4, i.e., nitrophenols and nitrobenzoic acids (□), have two nitro groups (Δ), contain a cyano substituent (O), and have two phenyl rings (◊). IGC_{50} is concentration that inhibits growth by 50 %.

The selection criteria were corroborated using results for the remaining data sets. Figure 2.1b and c shows the QSARs for the correlations between the decadic logarithm of the experimental toxicity of substituted mononitrobenzenes to *S. cerevisiae* and *D. magna*,

respectively, compared with the toxicity predicted based on log K_{ow}. After excluding outliers, improved correlations were obtained for both Figure 2.1b (r^2 = 0.85 instead of 0.02) and Figure 2.1c (r^2 = 0.83 instead of 0.08). For the remaining data sets, Table 2.1 displays statistical parameters for comparisons between correlations obtained in QSARs described with and without selection criteria. Notable improvements were with *Scenedesmus obliquus* (Eqn. 6), where r^2 increased from 0.02 to 0.81, and with *D. magna* (Eqn. 7), where r^2 increased from 0.02 to 0.83 after applying the selection criteria. Not all QSARs improved after applying selection criteria as shown by *C. vulgaris* (Eqn. 8) and *P. phosphoreum* (Eqn. 9), where r^2 slightly increased from 0.48 to 0.60 and from 0.04 to 0.13, respectively. A summary of the inclusion and exclusion criteria based on the substituents present on the substituted mononitrobenzenes is shown in Figure 2.2.

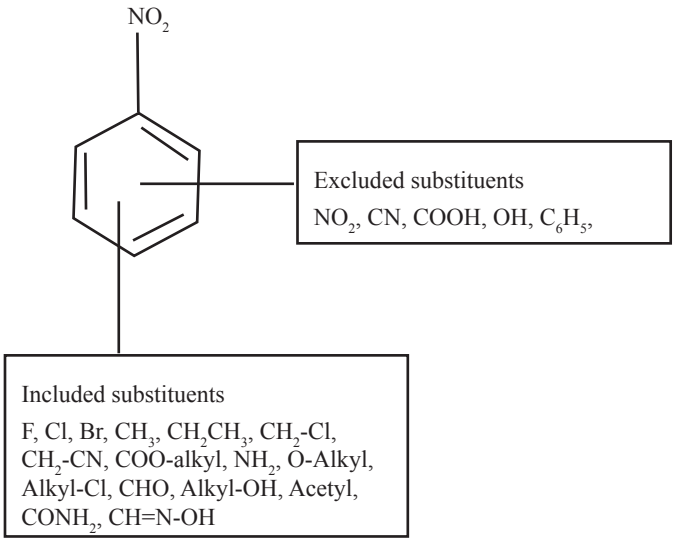


Fig. 2.2. Summary of included and excluded substituents for the substituted mononitrobenzene quantitative structure-activity relationships.

Two-parameter QSARs

In a next step, it was investigated whether the addition of a second parameter, E_{lumo}, in the QSARs for the substituted mononitrobenzenes meeting the selection criteria would result in improved correlations between experimental and predicted toxicity. This was in line with the observations of Cronin et al. (*9*), who demonstrated an improved QSAR relationship for nitrobenzenes after combining log K_{ow} and E_{lumo} as predictors. The use of E_{lumo} as an additional descriptor improved the QSARs only for algae species (Table 2.1) and was an

unnecessary descriptor for the other species. For *C. vulgaris*, combining log K_{ow} and E_{lumo} resulted in a better correlation (Eqn.10; r^2 = 0.85) than using log K_{ow} alone (Eqn. 8; r^2 = 0.60). A similar improvement was also noted for *C. pyrenoidosa* (Eqns. 11 and 12). Slight improvements were noted for the other algae genus, *Scenedesmus* (Eqns. 6 and 13–15).

L*og* K_{ow} *cutoff values*

The 10 non-polar narcotics models describing toxicity to different organisms that were compared to the log K_{ow}–based QSARs developed in this chapter are shown in Table 2.2. These QSARs were developed using toxicity data of simple alcohols, ketones, substituted aromatic hydrocarbons, ethers, and chlorinated hydrocarbons (*29*). The specific substituted mononitrobenzene QSARs of the present chapter to which they were compared are also indicated in this table. Where the non-polar narcotics QSAR was not available for an identical organism and endpoint, a comparison was made to the non-polar narcotics QSAR of a similar organism; for example, the QSAR for 96-h effective concentration to 50% of population (EC_{50}) to *C. carpio* (Table 2.1, Eqn. 19) was compared to the non-polar narcotics QSAR describing 96-h toxicity to the fish *Pimephales promelas* (Table 2.2, Eqn. 28). All non-polar narcotics QSARs shown in Table 2.2 had r^2 = 0.81. Figure 2.3 shows plots of log K_{ow} versus 48-h EC_{50} to *D. magna* for the QSAR for non-polar narcotics (---) toxicity (Table 2.2, Eqn. 26) and the one developed in this chapter (—) (Table 2.1, Eqn. 18). The two plots intersect at a log K_{ow} value of 4.08, which is then set as the upper limit for applying our QSAR for *D. magna*. Similar plots were made for the rest of the QSARs, and the log K_{ow} cutoff values are shown in Table 2.1. The log K_{ow} cutoff values ranged from 2.96 to > 10.

Extraction of substituted mononitrobenzenes meeting the selection criteria from the EINECS list

The EINECS list supplied by the European Chemicals Bureau contained 100 196 compounds with none of them containing a SMILES code. Cross-referencing the EINECS list with the SMILES-CAS database resulted in 54 277 compounds with SMILES codes. In Microsoft Excel 2000, based on selected aspects in the name, molecular formula, and SMILES code, compounds satisfying the set criteria were filtered out. Two attributes were most important in extracting the relevant substituted mononitrobenzenes from the EINECS list. These were the presence of N(=O) and c1 in the SMILES code, which indicate the presence of a nitro group and a ring, respectively. These attributes accounted for the extraction of 3 074 compounds out of 54 277.

Table 2.1. Quantitative structure-activity relationships (QSARs) for predicting toxic endpoints for various species exposed to substituted nitrobenzenes, using the logarithm of the octanol water partitioning coefficient (log K_{ow}) and energy of the lowest unoccupied molecular orbital (E_{lumo}) as descriptors. The numbers in parentheses show the values when no selection criteria[a] are used

	log toxicity (μM) = x(log K_{OW}) + y(E_{lumo}) + z											Ref for tox data
Eq.	Toxicity endpoint	x	y	z	r^2 [b]	n [c]	s [d]	F [e]	r^2_{int} [f]	log K_{ow} cut- off [g] ± SE	h_w [h]	
3	48 h IGC_{50} [i] *Tetrahymena pyriformis*	-0.54		3.62	0.81 (0.27)	33 (47)	0.20 (0.58)	138	0.78	5.51 ± 0.05	0.090	*(11)*
4	12 h EC_{50} [j] *Saccharomyces cerevisiae*	-0.66		6.03	0.85 (0.02	14 (24)	0.16 (0.60)	68	0.80	6.84 ± 0.08	0.213	*(10)*
5	21 d EC_{50} *Daphnia magna*	-0.67		3.40	0.83 (0.08)	15 (22)	0.13 (0.52)	62	0.72	4.20 ± 0.09	0.201	*(15)*
6	48 h EC_{50} *Scenedesmus obliquus*	-0.58		3.52	0.81 (0.02)	15 (25)	0.18 (0.53)	54	0.71	2.97 ± 0.08	0.201	*(16)*
7	21 d $LRCTr_m$ [k] *D. magna*	-1.04		4.07	0.83 (0.02)	15 (22)	0.20 (0.51)	62	0.79	3.72 ± 0.13	0.201	*(15)*
8	15 min EC_{50} *Chlorella vulgaris*	-0.90		4.82	0.60 (0.48)	10 (22)	0.66 (0.58)	12	ND[l]	> 10	ND	*(17)*
9	15 min EC_{50} *Photobacterium phosphoreum*	-0.48		2.75	0.13 (0.04)	15 (22)	0.53 (0.63)	2	ND	ND	ND	*(15)*
10	15 min EC_{50} *C. vulgaris*	-0.54	2.01	6.40	0.85 (0.77)	10 (22)	0.43 (0.40)	20	0.73	> 10	0.300	*(17)*
11	96 h EC_{50} *C. pyrenoidosa*	-0.59	1.45	4.94	0.80 (0.75)	15 (22)	0.26 (0.33)	24	0.60	> 10	0.201	*(15)*
12	96 h EC_{50} *C. pyrenoidosa*	-1.07		4.48	0.64 (0.02)	15 (22)	0.33 (0.63)	23	ND	> 10	0.201	*(15)*
13	48 h EC_{50} *S. obliquus*	-0.36	0.65	3.77	0.83 (0.20)	15 (25)	0.17 (0.49)	30	0.76	2.97 ± 0.08	0.201	*(16)*
14	96 h LC_{50} [m] *S. obliquus*	-0.57		3.49	0.80 (0.06)	13 (20)	0.19 (0.52)	43	0.70	2.96 ± 0.09	0.231	*(13)*
15	96 h LC_{50} *S. obliquus*	-0.35	0.65	3.75	0.82 (0.17)	13 (20)	0.18 (0.51)	24	0.71	2.96 ± 0.09	0.231	*(13)*
16	21 d $LRCT_{length}$ [n] *D. magna*	-0.81		3.59	0.65 (0.09)	15 (22)	0.25 (0.61)	24	0.52	4.41 ± 0.17	0.201	*(15)*
17	48 h IC50[o] *D. carinata*	-0.19		2.50	0.22 (0.01)	12 (21)	0.24 (0.45)	3	ND	ND	ND	*(13)*
18	48 h IC_{50} *D. magna*	-0.63		3.49	0.55 (0.06)	15 (22)	0.24 (0.44)	16	ND	4.08 ± 0.16	ND	*(15)*
19	96 h LC_{50} *Cyprinus carpio*	-0.52		3.40	0.62 (0.02)	11 (19)	0.28 (0.58)	15	ND	3.21 ±	ND	*(13)*
20	14 d LC_{50} *Poecilia reticulata*	-0.63		3.52	0.69 (0.00)	18 (26)	0.27 (0.63)	36	0.65	5.60 ± 0.11	0.168	*(14)*
21	40 h IGC_{50} *T. pyriformis*	-0.80		4.55	0.86 (0.20)	27 (39)	0.20 (0.61)	147	0.81	5.67 ± 0.07	0.111	*(9)*

[a] Neutral substituted mononitrobenzenes with no -OH, -COOH or -CN substituents attached directly to the ring. [b] Coefficient of determination. [c] Number of compounds used in QSAR. [d] Standard error of the estimate. [e] Variance ratio [f] Cross validated correlation coefficient. [g] Calculated by comparing own QSAR to similar nonpolar QSAR.
[h] Warning leverage = 3h where h is average leverage value for calibration compounds (*24*). [i] IGC_{50} = concentration inhibiting growth by 50 %. [j] EC_{50} = effective concentration to 50 % of population. [k] Lowest rejected concentration that significantly ($p < 0.01$) lowered the population growth constant (r_m) of daphnids (*15*). [l] Not determined – $q^2 < 0.65$; log K_{OW} cut-off not calculated where a similar baseline QSAR could not be found; h* not calculated where $q^2 < 0.5$. [m] LC_{50} = lethal concentration to 50 % of population. [n] Lowest rejected concentration that significantly ($p < 0.01$) lowered the mean length of daphnids (*15*). [o] IC_{50} = concentration immobilizing 50% of population

Table 2.2. Non-polar toxicity QSAR models for various species

Eqn.	Toxicity endpoint	log toxicity (μM) = x(log K_{ow}[a]) + y		Similar species QSAR compared to	n[b]	r^2[c]	s[d]	Ref
		x	y					
22	14 d LC_{50}[e] *Poecilia reticulata*	-0.87	4.87	24	50	0.976	0.31	*(29)*
23	48 h EC_{50}[f] *Tetrahymena pyriformis*	-0.80	5.20	3, 25	26	0.929	0.40	*(29)*
24	21 d LRCTr$_m$[g] *Daphnia magna*	-0.67	2.71	10	6	0.869	0.35	*(15)*
25	21 d $LRCT_{length}$[h] *D. magna*	-0.95	4.21	20	10	0.953	0.46	*(15)*
26	48 h EC_{50} *D. magna*	-0.95	4.81	22	17	0.990	0.21	*(29)*
27	96 h EC_{50} *S. capricornutum*	-1.00	4.77	14	10	0.929	0.17	*(29)*
28	96 h LC_{50} *Pimephales promephales*	-0.94	4.75	23	60	0.943	0.34	*(29)*
29	16 d EC_{50} *D. magna*	-0.64	3.27	7	5	0.990	0.08	*(15)*
30	12 h EC_{50} *Saccharomyces cerevisiae*	-0.79	6.91	5	19	0.980	0.08	*(10)*
31	15 min EC_{50} *Chlorella vulgaris*	-1.04	6.28	11	10	0.960	0.27	*(30)*

[a] Logarithm of the octanol/water partition coefficient.
[b] Number of compounds used in QSAR. [c] Coefficient of determination. [d] Standard error of the estimate.
[e] LC_{50} = lethal concentration to 50 % of population. [f] EC_{50} = effective concentration to 50 % of population.
[g] Lowest rejected concentration that significantly ($p < 0.01$) lowered the population growth constant (r_m) of daphnids (*15*).
[h] Lowest rejected concentration that significantly ($p < 0.01$) lowered the mean length of daphnids (*15*).

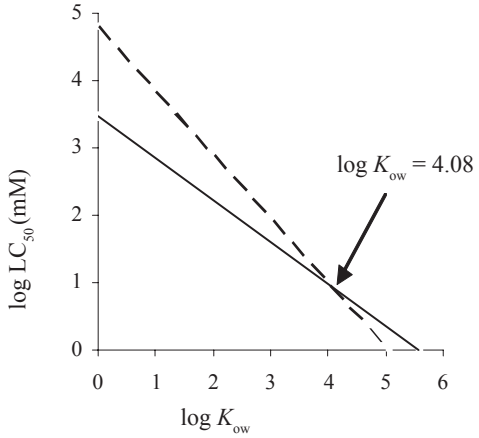


Fig. 2.3. QSARs for the correlation between the decadic logarithm of the experimental toxicity and log K_{ow} for substituted mononitrobenzenes (—) and for nonpolar narcotics (**---**) for 48 h log effective concentration to 50 % of population (EC_{50}) of *Daphnia magna.* The QSAR equations used for these plots are shown in Table 2.1 (Eqn. 22) and Table 2.2 (Eqn. 30). LC_{50} is lethal concentration to 50 % of population.

From the 45 919 compounds with no SMILES codes, 14 neutral, substituted mononitrobenzenes with no cyano substituent were extracted on the basis of the presence of carbon, nitrogen, and oxygen in the molecular formula. Full details of the extractions steps are shown as an appendix in Table S11. The final list of substituted mononitrobenzenes satisfying the conditions of the selection criteria for the newly defined QSARs contained 497 compounds and is shown as an appendix in Table S12.

DISCUSSION

Quality of developed QSAR models

The QSAR models were developed following the recommended guidelines set by the Organization for Economic Cooperation and Development (OECD) in November 2004 (*31*). The guidelines state that the following five major categories should be defined for each QSAR model: endpoint, unambiguous algorithm, internal performance and predictivity, domain of applicability, and a mechanistic interpretation, if possible (*31*). The biological data used were from well-standardized assays with clearly defined endpoints. The log K_{ow} estimation software models tested in this chapter are of good predictive quality, as plotting experimental against predicted log K_{ow} resulted in high coefficients of determination ($r^2 > 0.81$). Similar results were obtained previously for nitroaromatics by Petrauskas and Kolovanov (*21*). The differences that exist in the predictions made by the different software packages could be due mainly to the following aspects in the training set of the model: number and type of compounds, fragments, atoms, bond types, and correction factors. If more of the substituted nitrobenzenes in our QSARs are present in the training set of a particular model, then that model is likely to have a better predictive power. The prediction algorithms in IALogP and ALogPs are based on neural networks and a combination of electronic and topological attributes (*32*). CLogP, ACD/LogP, and K_{ow}WIN base their calculations on assigning values to different fragments in a molecule and then applying correction factors for the interactions between the fragments (*18, 21*). Spartan assigns contribution values to the different atoms in the molecule rather than to fragments (*33*). The performance of the software models could be further tested by comparing the correctness of the theoretical assumptions behind the calculation algorithms (*21*), which is not always easy with commercial products, as the algorithms are not always transparent. Furthermore, for untested compound classes, one should be aware that any estimation software may give inaccurate estimations, and this always has to be checked.

The chemical domain of the QSAR models was systematically developed. Outliers (dinitro-, cyano-, and ionizing nitrobenzenes) were omitted from the final QSARs because of their different uptake and reaction mechanisms. Substituted dinitrobenzenes have been shown previously to have enhanced toxicity when compared to substituted mononitrobenzenes, probably because of their fast reduction to the corresponding aryl-hydroxylamines (*9, 14, 15*), which justifies their exclusion. Compounds that can ionize at physiological pH (e.g., benzoic acids and nitrophenols) have different kinetics of uptake in the body than those that are neutral and may have additional toxicity by interfering with proton gradients during oxidative phosphorylation (*28*). With respect to 2-, 3-, and 4-cyanonitrobenzene, only

2-cyanonitrobenzene was an outlier. It could be due to experimental errors or to a mechanistic reason that is unclear at this moment.

The toxicity data used to establish the QSARs were from the same laboratory sources to avoid inter-laboratory variation (*28*). The statistical technique, linear regression, applied to analyze the relationship between toxicity and the physical-chemical descriptors is simple to use and sufficient, especially where two descriptors are involved.

Non-algal QSARs showed a high dependence on log K_{ow} for prediction of the toxicity of the substituted mononitrobenzenes, a relationship that has been demonstrated previously (*9, 14, 15*). Using the toxicity data of Deneer et al. (*14*), Banerjee and Williams (*34*) reported poor correlations between log K_{ow} and toxicity to the guppy *P. reticulata*. For the same data set, excluding substituted dinitrobenzenes from the QSAR analysis improved this poor correlation. With respect to the use of log K_{ow} as a toxicity descriptor, it is important to note that the relation between log K_{ow} and toxicity often remains quite empiric. Although log K_{ow} is often the most dominant descriptor for predicting the toxicity of certain compound classes to certain species, it is not universal, as was shown by its inapplicability to dinitro- and cyanonitrobenzenes in the present chapter. Furthermore, as octanol does not adequately represent the type of lipids found in membranes, other researchers have recommended replacing octanol with dimyristoyl phosphatidylcholine (DMPC) and thus log K_{ow} with log $K_{DPMC/water}$ (*35*).

In addition to log K_{ow} as a predictor, the inclusion of the reactivity parameter E_{lumo} improved the QSARs describing toxicity only to the algae *C. vulgaris*, *C. pyrenoidosa*, and *S. obliquus*. In the studies of Deneer et al. (*15*), a similar trend was observed after combining log K_{ow} and Hammett's reactivity parameter ($\Sigma\sigma^-$) when modeling the toxicity of substituted mononitrobenzenes to *C. pyrenoidosa*. They suggested that the mode of action or the toxicokinetic behavior of the substituted mononitrobenzenes in algae could be different from that of other organisms (*15*). Other QSARs of comparable quality describing the toxicity of substituted nitrobenzenes to various organisms have been reviewed by Katrizky et al. (*36*). Some of these QSARs have descriptors that are different from those used in this chapter. Roy and Ghosh (*37*) used the extended topochemical atom indices to model ($r^2 = 0.92$) toxicity to *T. pyriformis*. Although the extended topochemical indices show a significant potential for QSAR development, they still need extensive evaluation. In describing toxicity to *T. thermophila*, the best QSAR obtained by Xu et al. (*38*) used $\Sigma\sigma$- and the indicator variable *I* as descriptors ($r^2 = 0.852$). The indicator variable denotes the number and position of the nitro groups. However, no model validation was performed. Using the CODESSA-PRO software, Katrizky et al. (*36*) combined constitutional, topological, geometric, electrostatic,

and quantum chemical descriptors and obtained a decent QSAR ($r^2 = 0.72$) to describe toxicity to *T. pyriformis*. However, they combined toxicity data from two independent laboratories. The three QSARs mentioned previously could be improved by defining the applicability domain more clearly.

Toxicity prediction and priority setting

For new and existing chemicals, the OECD requires at least three basic acute ecotoxicity tests for risk assessment purposes. These tests describe acute toxicity to algae (72-h EC_{50} for growth), daphnids (48-h EC_{50} for immobilization), and fish (96-h median lethal concentration LC_{50}) (*39*). These tests evaluate toxicity at three trophic levels: primary producer, primary consumer, and top predator. In order to get a quick and broad overview of the toxicity of substituted mononitrobenzenes, for each organism in each of the trophic levels, the ''best'' QSAR was selected based on three criteria: $r^2_{int} \geq 0.5$, a long exposure period ($\geq$ 96 h where there was more than one exposure period , and the presence of a log K_{ow} cutoff value. Based on the three criteria, an advisory tool was developed and is shown in Figure 2.4.

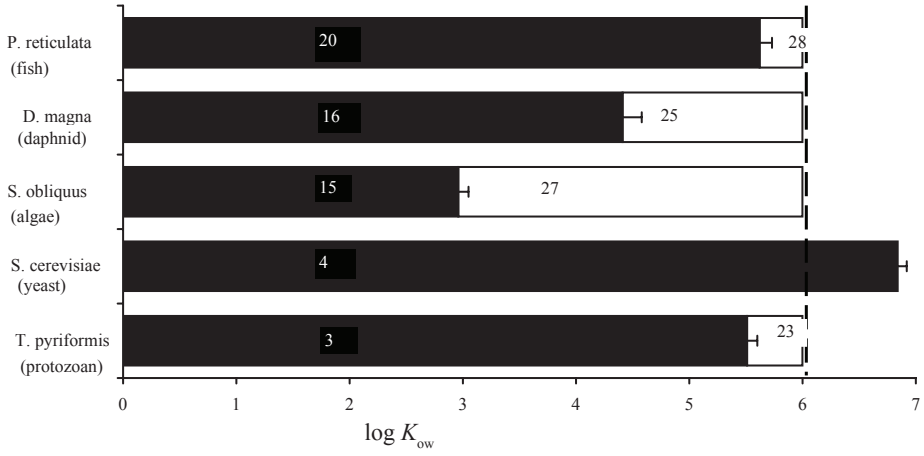


Fig. 2.4. Advisory tool for predicting acute toxicity of substituted mononitrobenzenes to five organisms. The black bars represent log K_{ow} ranges where the QSARs developed in this chapter are applicable. The white/open bars show log K_{ow} ranges where the non-polar narcotic QSARs yield the lowest effect levels. The numbers inside the bars refer to the appropriate QSAR models (Tables 2.1 and 2.2 respectively) to apply within that log K_{ow} range. Error bars show the error associated with the log K_{ow} cut-off points. The lower and upper (vertical dotted line) log K_{ow} limits are set at zero and six respectively.

Within the log K_{ow} ranges represented by the black bars, the QSAR models developed in this chapter are applicable. Above this range, shown in white/open bars, a risk manager should use the non-polar narcotics QSAR models. The lower and upper log K_{ow} limits of using Figure 2.4 are set at 0 and 6, respectively. Since log K_{ow} describes the kinetics of the uptake of chemicals from water, chemicals with log K_{ow} greater than 6 are generally taken up too slowly to show acute toxic effects, while those with log K_{ow} less than 0 would need unrealistically high concentrations to display toxicity (*40*).

Using the scheme in Figure 2.4, a risk manager selects the appropriate QSAR to use for prediction of toxicity based on the log K_{ow} of the compound. For example, a hypothetical substituted mononitrobenzene with log K_{ow} = 2 can have its toxicity to all species predicted by using the QSARs developed in this chapter. Another hypothetical substituted mononitrobenzene with log K_{ow} = 5 can have its toxicity predicted by using a combination of both non-polar narcotics QSARs and the QSARs developed in this chapter. Table 2.3 shows the predicted toxicities of five substituted mononitrobenzenes randomly selected from the list of 497 compounds meeting the selection criteria set. Table 2.3 shows that predicted toxicities increase with an increase in the log K_{ow} values. Predicted toxicities to *D. magna* after 21 d of exposure are consistently higher for all compounds when compared to the toxicity toward other organisms.

Although the advisory tool in Figure 2.4 shows QSARs for five specific species, it is a generally accepted approach in safety assessment in environmental toxicity that it can be used to predict the toxicity of other similar organisms as well (*41*), such as other fish, besides the specific species (*P. reticulata*) used for the QSAR. Although Figure 2.4 shows our recommended QSARs, other cross-validated QSARs ($r^2_{int} \geq 0.5$) shown in Table 2.1 can be used as well. The tool has two strong points. First, it quickly provides an overview of toxicity across different types of organisms, and, second, only high-quality cross-validated QSARs have been included. Whenever additional data become available, external validation and addition of QSARs for more animal species will help strengthen the tool further.

Table 2.3. The predicted acute toxicities of five substituted nitrobenzenes selected from the EINECS list

Compound	CAS[a] no.	log K_{ow}[b]	E_{lumo}[c] (eV)	Predicted[d] EC_{50}[e] (μM)				
				Poecilia reticulata (fish, 14 d)	*Daphnia magna (daphnid, 21 d)*	*Scenedesmus obliquus (algae, 96 h)*	*Saccharomyces cerevisiae(yeast, 12 h)*	*Tetrahymena pyriformis (protozoan, 48 h)*
N-(2-ethylhexyl)-2-nitroaniline	85117-98-2	5.76	-0.63	0.72	0.05	0.10	166.30	3.91
1,4-dibutoxy-2-nitrobenzene	135-15-9	5.07	-0.92	2.15	0.25	0.50	476.06	7.61
2,3,4,5-tetrachloronitrobenzene	2879-39-0	4.35	-1.83	6.08	1.17	2.63	1.4×10^3	18.65
1-chloro-2-fluoro-4-nitrobenzene	350-31-2	2.74	-1.6	62.41	23.51	56.17	1.7×10^4	139
2-nitrobenzyl alcohol	612-25-9	0.77	-1.15	1.1×10^3	928	539	3.4×10^5	1.6×10^3

[a] Chemical Abstracts Service.
[b] Logarithm of the octanol/water partition coefficient. Estimated using CLogP Version 4.0 (Biobyte, Claremont, CA, USA).
[c] Energy of the lowest unoccupied molecular orbital Estimated and Spartan 04 for Windows® Version 1.0.3 (Wavefun, Irvine, CA, USA;)
[d] Toxicity predicted according to the advisory tool shown in Figure 2.4. The type of organism and period of exposure are shown underneath the name of the organism.
[e] EC_{50} = effective concentration to 50 % of population.

Future perspectives

In case industry considers producing a new substituted mononitrobenzene that fits the selection criteria, using the methodology established here, it is possible to estimate the toxicological effects of this new compound with no additional toxicological or animal testing. Although the QSARs presented here refer to 497 out of the more than 100 000 EINECS chemicals (i.e., 0.5 %), QSARs will often be restricted to specific chemical classes, as it is unlikely that one QSAR model will ever describe the full set of EINECS chemicals. Considering this, it is a step forward that 0.5 % of the EINECS chemicals can be included in the QSARs described in the present chapter.

Appendices

Supplementary information is available in 11 tables and is available online (*42*). Table S1: log K_{ow} and E_{lumo} values used to develop the QSARs in Table 2.1; Tables S2 to S9: experimental toxicity data sets; Table S11: systematic procedure for extracting the compounds from the EINECS list satisfying the selection criteria of our QSARs; and Table S12: the list of the extracted 497 EINECS compounds.

Acknowledgements

The Netherlands Organization for Health Research and Development (ZonMw) funded this research under the theme ‘‘Alternatives to Animal Experimentation.’’ The authors also wish to thank Remi Allanou (European Chemicals Bureau) for supplying the EINECS list and Bill Meylan (Syracuse Research, NY, USA) for providing the SMILES-CAS database and the reviewers for their constructive comments.

References

(1) EU. (2001) White Paper: Strategy for a future chemicals policy. Commission of the European Committees, Brussels, Belgium. COM(2001) 88 final., pp 1–32.

(2) Worth, A. P., Van Leeuwen, C. J. and Hartung, T. (2004) The prospects for using (Q)SARs in a changing political environment - High expectations and a key role for the European Commission's joint research centre. *SAR QSAR Environ Res 15*, 331.

(3) NHC. (2001) Toxicity testing: a more efficient approach. Publication no. 2001/24E. The Hague, The Netherlands.

(4) USEPA. (2002) Endocrine Disruptor Screening Program, Proposed Chemical Selection Approach for Initial Round of Screening; Request for Comment. FRL-7286-6, Washington DC.

(5) Hansch, C. and Fujita, T. (1964) A method for the correlation of biological activity and chemical structure. *J. Amer. Chem. Soc. 86*, 1616 - 1626.

(6) Abraham, M. H., Chadha, H. S., Martins, F., Mitchell, R. C., Bradbury, M. W. and Gratton, J. A. (1999) Hydrogen bonding part 46: A review of the correlation and prediction of transport properties by an LFER method: Physicochemical properties, brain penetration and skin permeability. *Pesticide Science 55*, 78.

(7) Soffers, A. E. M. F., Boersma, M. G., Vaes, W. H. J., Vervoort, J., Tyrakowska, B., Hermens, J. L. M. and Rietjens, I. M. C. M. (2001) Computer-modeling-based SARs for analyzing experimental data on biotransformation and toxicity. *Toxicol in Vitro 15*, 539-551.

(8) Freidig, A. P. and Hermens, J. L. (2000) Narcosis and chemical reactivity QSARs for acute fish toxicity. *Quant. Struct. Act. Relat. 19*, 547 - 553.

(9) Cronin, M. T., Gregory, B. W. and Schultz, T. W. (1998) Quantitative structure-activity analyses of nitrobenzene toxicity to *Tetrahymena pyriformis*. *Chem. Res. Toxicol. 11*, 902-908.

(10) Wang, X., Yin, C. and Wang, L. (2002) Structure-activity relationships and response-surface analysis of nitroaromatics toxicity to the yeast (Saccharomyces cerevisiae). *Chemosphere 46*, 1045-1051.

(11) Bohac, M., Loeprecht, B., Damborsky, J. and Schüürmann, G. (2002) Impact of orthogonal signal correction (OSC) on the predictive ability of CoMFA models for the ciliate toxicity of nitrobenzenes. *Quant. Struct. Act. Relat. 21*, 3 - 11.

(12) O'Brien, P. J., Wong, W. C., Silva, J. and Khan, S. (1990) Toxicity of nitrobenzene compounds towards isolated hepatocytes: dependence on reduction potential. *Xenobiotica; The Fate Of Foreign Compounds In Biological Systems 20*, 945-955.

(13) Zhao, Y.-H., Yuan, X., Ji, G.-D., Sheng, L.-X. and Wang, L.-S. (1997) Quantitative structure-activity relationships of nitroaromatic compounds to four aquatic organisms. *Chemosphere 34*, 1837-1844.

(14) Deneer, J. W., Sinnige, T. L., Seinen, W. and Hermens, J. L. M. (1987) Quantitative structure-activity relationships for the toxicity and bioconcentration factor of nitrobenzene derivatives towards the guppy (Poecilia reticulata). *Aquat. Toxicol. 10*, 115-129.

(15) Deneer, J. W., van Leeuwen, C. J., Seinen, W., Maas-Diepeveen, J. L. and Hermens, J. L. M. (1989) QSAR study of the toxicity of nitrobenzene derivatives towards *Daphnia magna, Chlorella pyrenoidosa* and *Photobacterium phosphoreum*. *Aquat. Toxicol. 15*, 83-98.

(16) Lu, G.-H., Yuan, X. and Zhao, Y.-H. (2001) QSAR study on the toxicity of substituted benzenes to the algae (*Scenedesmus obliquus*). *Chemosphere 44*, 437-440.

(17) Netzeva, T. I., Dearden, J. C., Edwards, R., Worgan, A. D. P. and Cronin, M. T. D. (2004) QSAR analysis of the toxicity of aromatic compounds to *Chlorella vulgaris* in a novel short-term assay. *J. Chem. Inf. Comp. Sci. 44*, 258-265.

(18) Leo, A. (1993) Calculating log Poct from structures. *Chem. Rev. 93*, 1281-1306.
(19) Tetko, I. V. and Tanchuk, V. Y. (2002) Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *J. Chem Info Comp Sci 42*, 1136-1145.
(20) Meylan, W. M. and Howard, P. H. (1995) Atom/fragment contribution method for estimating octanol-water partition coefficients. *J. Pharm. Sci. 84*, 83-92.
(21) Petrauskas, A. A. and Kolovanov, E. A. (2000) ACD/Log P method description. *Perspect. Drug Discovery Des. 19*, 99-116.
(22) Weininger, D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules *J. Chem. Inf. Comput. Sci. 28*, 31-36.
(23) Dewar, M., Zoebisch, E., Healy, E. and Stewart, J. (1985) AM1: A new general purpose quantum mechanical molecular model. *Amer. Chem. Soc. 107*, 3902-3909.
(24) Tropsha, A., Gramatica, P. and Gombar, V. K. (2003) The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR and Combinatorial Science 22*, 69-77.
(25) Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., McDowell, R. M. and Gramatica, P. (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ. Health Perspect. 111*, 1361–1375.
(26) Diamantidis, N. A., Karlis, D. and Giakoumakis, E. A. (2000) Unsupervised stratification of cross-validation for accuracy estimation. *Artificial Intelligence 116*, 1-16.
(27) Khadikar, P. V., Karmarkar, S., Singh, S. and Shrivastava, A. (2002) Use of the PI index in predicting toxicity of nitrobenzene derivatives. *Bioorg Med Chem 10*, 3163-3170.
(28) Cronin, M. T. D. and Schultz, T. W. (2003) Pitfalls in QSAR. *Journal of Molecular Structure: THEOCHEM 622*, 39-51.
(29) Van Leeuwen, C. J., Van der Zandt, P. T. J., Aldenberg, T., Verhaar, H. J. M. and Hermens, J. L. M. (1992) Application of QSARs, extrapolation and equilibrium partitioning in aquatic effects assessment. I. Narcotic industrial pollutants. *Environ. Toxicol. Chem. 11*, 267-282.
(30) Worgan, A. D. P., Dearden, J. C., Edwards, R., Netzeva, T. I. and Cronin, M. T. D. (2003) Evaluation of a novel short-term algal toxicity assay by the development of QSARs and inter-species relationships for narcotic chemicals. *QSAR & Combinat. Sci. 22*, 204-209.
(31) OECD. (2004) The report from the expert group on (Quantitative) Structure-Activity Relationship ((Q)SARs) on the principles for the validation of (Q)SARS. ENV/JM/MONO(2004)24. , In *OECD Series on testing and assessment: number 49*.
(32) Hall, L. H. and Kier, L. B. (1995) Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *J. Chem. Inf. Comp. Sci. 35*, 1039-1045.
(33) Ghose, A., Pritchett, A. and Crippen, G. (1988) Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships III: Modeling Hydrophobic Interactions. *J. Comp. Chem. 9*, 80 - 90.
(34) Banerjee, S. and Williams, C. L. (1993) Compatibility of functional groups in K(ow)-based QSARs: Application to nitro compounds. *Environ Toxicol Chem 12*, 1847-1849.
(35) Vaes, W. H., Ramos, E. U., Verhaar, H. J., Cramer, C. J. and Hermens, J. L. (1998) Understanding and estimating membrane/water partition coefficients: approaches to derive quantitative structure property relationships. *Chem Res Toxicol 11*, 847-854.
(36) Katritzky, A. R., Oliferenko, P., Oliferenko, A., Lomaka, A. and Karelson, M. (2003) Nitrobenzene toxicity: QSAR correlations and mechanistic interpretations. *Journal of Physical Organic Chemistry 16*, 811 - 817.
(37) Roy, K. and Ghosh, G. (2004) QSTR with extended topochemical atom indices. 4. Modeling of the acute toxicity of phenylsulfonyl carboxylates to Vibrio fischeri using principal component factor analysis and principal component regression analysis. *QSAR and Combinatorial*

Science 23, 526.

(38) Xu, J. B., Jing, T. S., Pauli, W. and Berger, S. (2002) Quantitative structure-activity relationships for the toxicity of nitrobenzenes to Tetrahymena thermophila. *Journal of Environmental Science and Health - Part A Toxic/Hazardous Substances and Environmental Engineering 37*, 563.

(39) OECD. (2001) Guidance document on the use of the harmonized system for the classification of chemicals which are hazardous to the aquatic environment. Technical Report ENV/JM/MOMO (2001)8. Paris, France.

(40) Verhaar, H. J. M., van Leeuwen, C. J. and Hermens, J. L. M. (1992) Classifying environmental pollutants. *Chemosphere 25*, 471-491.

(41) EC. (2002) Guidance document on aquatic ecology. Sanco/3268/2001 Technical Report. Brussels, Belgium.

(42) Zvinavashe, E. (2006) Supplementary information available online at http://www.setacjournals.org/archive/1552-8618/25/9/extref/i1552-8618-25-9-2313-ETC-25-09-001.pdf.

QSAR models for predicting *in vivo* aquatic toxicity of chlorinated alkanes to fish

Elton Zvinavashe, Hans van den Berg, Ans E.M.F. Soffers, Jacques Vervoort, Andreas Freidig, Albertinka J. Murk, and Ivonne M.C.M. Rietjens

Chemical Research in Toxicology (2008) 21: 739–745

Abstract

Quantitative structure–activity relationship (QSAR) models are expected to play a crucial role in reducing the number of animals to be used for toxicity testing resulting from the adoption of the new European Union chemical control system called Registration, Evaluation, and Authorization of Chemicals (REACH). The objective of the present study was to generate *in vitro* acute toxicity data that could be used to develop a QSAR model to describe *in vivo* acute toxicity of chlorinated alkanes. Cytotoxicity of a series of chlorinated alkanes to Chinese hamster ovary (CHO) cells was observed at concentrations similar to those that have been shown previously to be toxic to fish. Strong correlations exist between the acute *in vitro* toxicity of the chlorinated alkanes and (i) hydrophobicity [modeled by the calculated log K_{ow} (octanol–water partition coefficient); $r^2 = 0.88$ and $r^2_{int} = 0.85$] and (ii) *in vivo* acute toxicity to fish ($r^2 = 0.76$). A QSAR model has been developed to predict *in vivo* acute toxicity to fish based on the *in vitro* data and even on *in silico* log K_{ow} data only. The developed QSAR model is applicable to chlorinated alkanes with up to 10 carbon atoms, up to eight chlorine atoms, and log K_{ow} values lying within the range from 1.71 to 5.70. Out of the 100 196 compounds on the European Inventory of Existing Chemicals (EINECS), our QSAR model covers 77 (0.1%) of them. Our findings demonstrate that *in vitro* experiments and even *in silico* calculations can replace animal experiments in the prediction of the acute toxicity of chlorinated alkanes.

Introduction

The European Inventory of Existing Chemicals (EINECS) (*1*) contains over 100 000 chemicals that are marketed within Europe for industrial or consumer needs. For most of these chemicals, there is insufficient (eco)toxicological information on their hazardous properties. To close these existing toxicity data gaps, the European Union parliament recently adopted a new chemical control system called Registration, Evaluation and Authorization of Chemicals (REACH) (*1*). One of the aims of REACH is to improve the protection of human health and the environment by requiring industry to provide toxicity information for the chemicals that they manufacture or distribute. There is currently an ongoing debate about the potentially large number of animals that have to be used for experimental toxicity testing as an outcome of REACH. Within REACH, however, there is a provision to use, among others, sufficiently validated computational prediction models based on quantitative structure–activity relationships (QSAR) to fill in the toxicity data gaps and thus save time and costs, reducing the number of experimental animals used. To increase the acceptability of QSAR models within REACH, guidelines for QSAR model development and validation proposed by the Organization for Economic Cooperation and Development (OECD) (*2*) are now widely accepted.

Chlorinated alkanes are an important group of chemicals on the EINECS list with widespread use, large production volumes, and thus a large potential for environmental pollution, and they are the focus of this chapter. Chlorinated *n*-alkanes are built from straight chains of carbon and hydrogen with varying numbers of hydrogen atoms replaced by chlorine atoms. The introduction of chlorine atoms into the hydrocarbon chain alters properties such as solubility, density, volatility, and toxicity (*3*). Some of these changes confer improvements that make the compounds useful commercially, but these changes can also make them more toxic. Chlorinated *n*-alkanes are broadly divided into two main groups depending on the number of carbon atoms present: lower chlorinated alkanes (LCA; C_1–C_9) and polychlorinated *n*-alkanes (PCA; C_{10}–C_{30}) (*4*). Mixtures of commercial PCAs, known as chlorinated paraffins, are divided into three groups: short-chain (C_{10}–C_{13}), medium-chain (C_{14}–C_{17}), and long-chain (C_{18}–C_{30}) paraffins with chlorine contents varying from 35–70 % by weight. The LCAs are widely used as industrial and household solvents, fumigants, and intermediates in chemical synthesis (*5*). The PCAs are often used as lubricating additives, adhesives, and flame retardants in rubber and textiles. The annual production volume of PCAs is greater than 300 kilotons (*4*). For the PCAs, toxicity is believed to decrease from the short to the long-chain PCAs, due to a decrease in solubility (*4*). It has been suggested

that the short-chain PCAs should be included in the list of persistent organic pollutants (*6*). They have a high bioaccumulation potential due to their high log K_{ow} (octanol–water partition coefficient) values, are persistent in the environment due to their resistance to degradation, and thus have a potential for long-range environmental transport. Short-chain PCAs are known to be highly toxic to aquatic organisms (*6*).

Despite their widespread use and presence, the amount of toxicological data on PCAs is rather limited. Because they are produced via free radical chlorination, a single PCA preparation can consist of many different congeners with a wide range of physicochemical properties (*4*). This presents problems in attempting to estimate the toxicity of these preparations as the toxicity of individual compounds cannot be identified. Using synthesized PCA congeners, Fisk and co-workers described their bioaccumulation in rainbow trout (*Onchorhynchus mykiss*) (*7*) and toxicity to Japanese medaka (*Oryzias latipes*) embryos (*4*).

For the LCAs, several studies describe their acute toxicity in literature. Crebelli and co-workers used electrophilicity descriptors to describe their aneugenic activity to the mold *Aspergillus nidulans (8)*. The acute toxicity of LCAs to the protozoan *Tetrahymena pyriformis (9)*, the marine bacterium *Photobacterium phosphoreum* in the Microtox test (*10*), the fathead minnow *Pimephales promelas (11)* the guppy *Poecilia reticulate (12)*, and HeLa cells (*13*) was determined and related to the hydrophobicity of the compounds. However, in most of these studies, the number of chlorinated alkanes tested was either too small to be used for QSAR modeling or was for a small range of carbon chain lengths [e.g., C_1 to C_5 as in studies by Könemann (*12*)]. As far as we are aware, there exists no QSAR model to predict the *in vivo* acute toxicity of chlorinated alkanes. The objective of the present study was to generate *in vitro* acute toxicity data that could be used to develop a QSAR model to describe *in vivo* acute toxicity of chlorinated alkanes. Toxicity tests were performed for a large set of chlorinated alkanes across a wide range of hydrophobicity values and carbon chain lengths (C_1–C_{10}). The *in vitro* toxicity data were used to develop a validated QSAR model with defined applicability limits following OECD guidelines. The *in vitro* toxicity data were further compared to *in vivo* toxicity data for fish, and a prediction model for *in vivo* toxicity was developed using the *in vitro* data. Finally, an estimate was made of the number of EINECS compounds for which the QSAR model can make accurate predictions.

MATERIALS AND METHODS

Materials

Unless otherwise indicated, all chemicals were obtained from Sigma-Aldrich (Zwijndrecht, The Netherlands) and were at least 98% pure. Stocks of chlorinated alkanes were prepared in spectrophotometric grade dimethyl sulfoxide (DMSO) obtained from Acros Organics (Geel, Belgium). Phosphate-buffered saline (PBS), Hank's balanced salt solution (HBSS), Dulbecco modified Eagle's medium (DMEM)/Ham's F12, fetal calf serum (FCS), and Trypsin-EDTA were supplied by Gibco-Invitrogen (Breda, The Netherlands). Chinese hamster ovary (CHO) wild-type cells were sourced from the American Type Culture Collection (Manassas, VA). Cell culture flasks (75 cm^2) were supplied by Corning Inc. (Corning, NY), and culture plates (24- and 96-wells) were provided by Greiner Bio-one (Alphen aan de Rijn, The Netherlands).

Cell culture

The CHO cells were grown in 75 cm^2 culture flasks and maintained in a humidified incubator at 37 ^oC, 95 % air/5 % CO_2 in DMEM/F12 medium supplemented with 10 % FCS. Once every three days, the cells were rinsed with HBSS, trypsinized, and then resuspended and cultivated in fresh culture medium. Cells from culture flasks with confluency of at least 90 % were used for the cytotoxicity assay.

Cytotoxicity assay

The cytotoxicity of the chlorinated alkanes to CHO cells was determined in triplicate in 96-well culture plates using the MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] assay following procedures described previously (*14*) with some modifications. Where appropriate, dilution of the test compounds was done in 24-well culture plates. Two methods of exposure were compared, direct and premix (the most commonly used exposure method during *in vitro* testing). For direct exposure, there was no prior dilution of the test compound in the dilution plate; the test compound solution in DMSO was added straight into the medium of the culture plate. For premix exposure, there was prior dilution of the test compound solution in DMSO using culture medium, and this culture medium containing the test compound at the desired concentration was added to the cells in the culture plate. For both methods, 100 µL of CHO cell suspension was seeded into each of the inner wells of the culture plate at a final concentration of 3×10^5 cells/ml, with 200 µL of HBSS added to the outer wells. The plate was incubated for 24 h to allow the cells to attach. For the direct

exposure, 100 µL of culture medium with 10 % FCS (37 °C) was then added to the inner wells followed by addition of 2 µL of each test compound solution at various concentrations. For the premix exposure, 50 times concentrated solutions of the test compounds in DMSO were diluted 50 times in culture medium with 10 % FCS in the dilution plate, and then, 100 µL of these medium samples was transferred to the culture plate. For both methods in each independent experiment, (i) all concentrations were tested in 6-fold and (ii) two sets of controls were used, one with DMSO and another with culture medium without the test compounds or DMSO. Upon addition of the test compounds, the plates were shaken at 600 rpm on an orbital shaker (Incubator 1000, Hieroglyph Instruments, Germany) for 5 min, followed by 21 h of incubation in the humidified incubator at 37 °C. Then the MTT reagent was added to a final concentration of 0.5 mg/ml, and incubation continued for a further 3 h, for a total incubation time of 24 h. After this, the culture medium was removed with a vacuum pump and 100 µL of DMSO was added to lyze the cells. The plates were shaken for 5 min to dissolve the formazan crystals formed after reduction of MTT. Subsequently, two absorbance readings were recorded as follows: A_{562} for the color of the formazan crystals and A_{620} for cell debris and other non-specific absorbance.

Calculation of EC_{50} values.

The A_{620} values were subtracted from the A_{562} values, and the result was expressed as a percentage of the response of the DMSO control. The EC_{50} values of the chlorinated alkanes were calculated using a Microsoft Excel plug-in, Life Sciences Workbench (LSW) Data Analysis Toolbox Version 1.1.1. (MDL Information Systems, CA) with the general sigmoidal curve with Hill slope (*a-d*) chosen as the best fit model.

Calculation of theoretical descriptors.

Hydrophobicity of the chlorinated alkanes was modeled using log K_{ow} values calculated using the software CLogP version 4.0 (Biobyte, Claremont, CA) (*15*) as described previously (*16*). Briefly, the structure of each molecule was entered into CLogP as a simplified molecular input line entry system (SMILES) code. The SMILES codes were obtained from the SMILES-CAS database (Syracuse Research, Syracuse, NY). Solubility values were calculated using ACD/Laboratories version 8.14 for Solaris (Advanced Chemistry Development, Toronto, ON, Canada).

QSAR modeling

The Statistical Package for Social Scientists (SPSS) version 13 for Windows (SPSS,

Chicago, IL) was used to analyze the QSARs with log EC_{50} as the dependent variable and log K_{ow} as the independent variable. The quality of the QSAR model was characterized by the number of compounds used in the study (n), coefficient of determination (r^2), standard error of the estimate (s), variance ratio (F), the internally cross-validated coefficient of determination (r^2_{int}), and the externally validated coefficient of determination (r^2_{ext}). Internal validation of the QSAR model was performed using the leave-out-many cross-validation method, with 20% of the calibration compounds left out at each step as described previously (*16*). The external performance of the QSAR model was evaluated by testing five additional compounds that fit into the applicability domain of the model and then comparing the predicted and experimental toxicity values. The calculation of r^2_{ext} was performed according to the formula:

$$r^2_{ext} = 1\text{-PRESS/SSD} \qquad \text{(Eq. 1)}$$

where PRESS (predicted sum of squares) is the sum of the squared differences between the predicted and the experimental toxicity values for each molecule in the validation set, and SSD is the sum of the squared deviations between the experimental toxicity values for each molecule in the validation set and the mean experimental toxicity values of the training set (*17*).

RESULTS

Twenty-one LCAs were tested in the MTT assay with CHO cells, and these model compounds of the present study are listed in Table 3.1, together with their estimated log K_{ow}, water solubility, and the experimental EC_{50} values obtained. The MTT cytotoxicity results were obtained using the direct exposure method. When comparing the cytotoxicity of the same concentration of LCAs after premix and direct exposure of CHO cells, major differences were found. After direct exposure to increasing concentrations of 1-chlorononane, for example, a clear dose dependent decrease in cell viability was observed (Figure 3.1). After premix exposure, however, none of the tested concentrations induced any cytotoxicity. Microscopic examination of the 96-well plate 10 min after exposure revealed no local cytotoxicity due to a possible temporary high concentration of DMSO when the direct exposure method was used. Because of the lack of a dose-response relationship with the premix method, further experiments were conducted using the direct exposure method to determine the EC_{50} values reported in Table 3.1.

Table 3.1: Chlorinated alkanes present in the training set, their Chemical Abstract Service (CAS) numbers, octanol/water partitioning coefficient (K_{ow}), water solubility, *in vitro* and *in vivo* EC_{50} values. The EC_{50} values in the present study were determined using an MTT test in Chinese Hamster Ovary (CHO) cells.

No.	Compound name	CAS no.	log K_{ow}[a]	log H_2O solubility[b]	*in vitro* log EC_{50} (μM) MTT assay			*in vivo* log EC_{50} (μM)		
				at 25°C (μM)	CHO cells[c] 24 h	Rat primary hepatocytes[d] 2 h	HeLa cells[e] 72 h	*Poecilia reticulata*[f] 7 d	*Phosphobacterium phosphoreum*[g], 5 min	*Pimephales promelas*[h] 96 h
1	1,3-dichloropropane	142-28-9	1.71	4.11	2.99			2.87	3.02	3.06
2	1,2-dichloropropane	78-87-5	1.99	4.46	3.03			3.01		3.09
3	1,1,2-trichloroethane	79-00-5	2.05	4.42	3.04	3.4		2.85		2.79
4	1,2-dichlorobutane	616-21-7	2.52	3.87	2.61			2.39[j]		
5	1,1,2,2-tetrachloroethane	79-34-5	2.64	3.849	2.44	3.24		2.34		2.09
6	1-chloro-2,2-dimethylpropane	753-89-9	2.79	3.76	2.3			2.11[j]		
7	Carbon tetrachloride	56-23-5	2.88	3.72	2.4	2.89	1	2.20[j]		
8	1-chloro-2-methylbutane	616-13-7	2.92	3.54	2.3			2.11[j]		
9	1-chloropentane	543-59-9	3.05	3.34	2.24			2.05[j]	2.55	
10	1,6-dichlorohexane	2163-00-0	3.29	2.56	2.02			1.85[j]		
11	1-chlorohexane	544-10-5	3.58	2.76	2.24			2.05[j]		
12	1,1-dichloro-3,3-dimethylbutane	6130-96-7	3.63	2.94	1.67			1.52[j]		
13	1-chloroheptane	629-06-1	4.11	2.18	1.73			1.58[j]		
14	1,8-dichlorooctane	2162-99-4	4.35	1.48	1.72			1.57[j]		
15	1-chlorooctane	111-85-3	4.64	1.61	1.43			1.30[j]		
16	1,9-dichlorononane	821-99-8	4.88	0.94	1.78			1.63[j]		
17	1-chlorononane	2473-01-0	5.17	1.04	1.35			1.23[j]		
18	1-chlorodecane	1002-69-3	5.7	0.52	1.44			1.31[j]		
19	1,10-dichlorodecane	2162-98-3	5.41	0.88	NTAS[i]					
20	1-chlorododecane	112-52-7	6.76	0.18	NTAS					
21	1-chlorotetradecane	2425-54-9	7.81	1.38	NTAS					

[a] calculated using ClogP version 4.0 (Biobyte, Claremont, CA, USA), [b] calculated using ACD/Labs version 8.14 for Solaris (ACD, Toronto, ON, Canada), [c-h] experimental toxicity data from [c]this thesis, [d](*18*), [e](*13*), [f](*12*), [g](*10*), and [h](*11*), [i] not toxic at saturation, [j] EC_{50} values predicted using Eq. 3

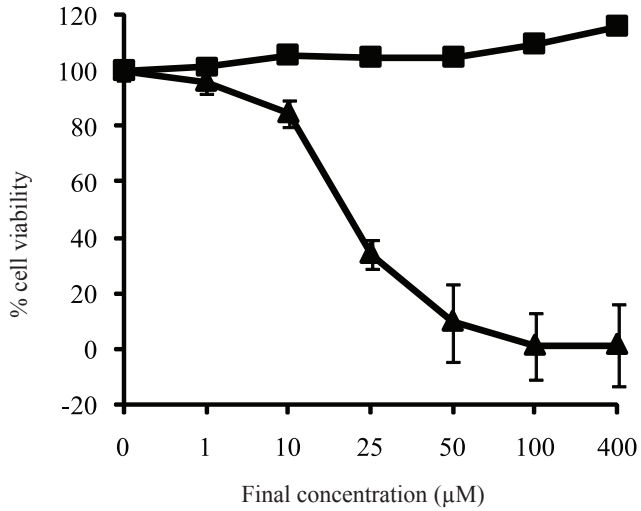


Fig. 3.1: Comparison of the cytotoxicity of 1-chlorononane to Chinese hamster ovary cells using direct (▲) and premix (■) methods of exposure in the MTT assay. The points on the graph represent the mean ± standard error of three experiments.

Influence of chain length and extent of chlorination on toxicity.

With an increase in chain length for single chlorinated compounds from 1-chloropentane, 1-chlorohexane, 1- chloroheptane, and 1-chlorooctane to 1-chlorononane, the cytotoxicity increased, with 1-chlorodecane showing a deviation (Table 3.1). For compounds with the same hydrocarbon backbone, an additional chlorine atom was associated with a decrease in toxicity. This is shown for example by 1,6-dichlorohexane vs 1-chlorohexane and 1,8-dichlorooctane vs. 1-chlorooctane. The short-chain PCAs 1,10-dichlorodecane, 1-chlorododecane, and 1-chlorotetradecane did not show cytotoxicity up to the maximum soluble concentrations tested; therefore, they were excluded from the modeling process.

QSAR modeling

An investigation of the relationship between log K_{ow} and the experimental toxicity data from the MTT test revealed a good correlation (Figure 3.2) and can be described by the following equation:

$$\log EC_{50}\,(\mu M) = -0.45(\log K_{ow}) + 3.69 \qquad \text{(Eq. 2)}$$

where $n = 18$, $s = 0.19$, $F = 120$, $r^2 = 0.88$, and $r^2_{int} = 0.85$.

An increase in log K_{ow} was associated with an increase in toxicity (Figure 3.2). However, with increasing log K_{ow} values, also the solubility decreased until above log K_{ow} = 4.53, where the EC_{50} becomes higher than the calculated water solubility (Figure 3.2). However, 1-chlorononane (log K_{ow} = 5.17) was still able to show toxicity above this cut-off value.

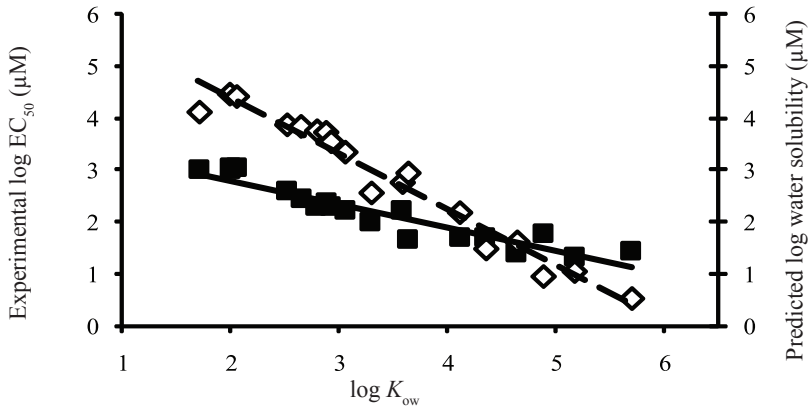


Fig. 3.2: Quantitative structure-activity relationship for the correlation between the decadic logarithm of the 24 h *in vitro* toxicity (EC_{50}) of chlorinated alkanes (C_1 – C_{10}) to Chinese hamster ovary cells in the MTT assay and the toxicity predicted based on log K_{ow} (■). The dotted line (◊) shows a plot of log K_{ow} versus water solubility of the chlorinated alkanes (n = 18, r^2 = 0.96). The two plots cross where log K_{ow} = 4.53.

Applicability domain of the QSAR model

The range of compounds for which the model is valid (applicability domain) was determined by taking into account the minimum and maximum values of both the (i) carbon chain length (C_1–C_{10}), (ii) the chlorine atoms (Cl_1–Cl_8), and (iii) the log K_{ow} values (1.71–5.70) of the compounds included in the training set. For models within a one-dimensional descriptor space, the interpolation region is simply taken as the interval between the minimum and the maximum values of the training data set (*19*). The theoretical toxicity cut-off limit of log K_{ow} = 4.53 was not taken into account here as compounds with higher log K_{ow} values were able to show toxicity. Taking these criteria into consideration, our QSAR model is thus applicable to chlorinated alkanes with up to 10 carbon and eight chlorine atoms and log K_{ow} values between 1.71 and 5.70. These limits were used as selection criteria to extract compounds with similar properties from the EINECS list. Using procedures described previously (*16*), molecular formula, name, and SMILES codes of the EINECS compounds were used as filtering criteria in Microsoft Excel 2003 to extract 59 compounds satisfying the selection criteria. The 59 compounds and their predicted EC_{50} values are shown in Table 3.2.

This implies that including the 18 compounds used to develop the MTT assay-based *in vitro* QSAR in CHO cells, our QSAR model covers 77 (0.1 %) of the EINECS list compounds.

External validation of QSAR model

From Table 3.2, five compounds, with log K_{ow} values within the applicability domain of the training set were selected for additional testing to externally validate our QSAR model. Suitable test concentrations were selected based on the predicted *in vitro* toxicity values, thereby skipping range-finding tests. The experimental toxicity values, shown in Table 3.2, correlated well with predicted values ($r^2_{ext} = 0.74$).

In vitro to in vivo correlation.

In a final step, it was investigated whether the acute *in vitro* toxicity data generated in the present study could be used to build a model for making acute *in vivo* toxicity predictions for fish. Tables 3.1 and 3.2 show the acute *in vivo* toxicity data for the fish, *P. reticulata (12)*, for the compounds that were also tested in the present study. A correlation of the *in vitro* and *in vivo* toxicity data is shown in Figure 3.3 and can be described by the following equation:

$$\textit{in vivo}\ \log LC_{50}\,(\mu M) = 0.93[\textit{in vitro}\ \log EC_{50}\,(\mu M)] - 0.02 \qquad \text{(Eq. 3)}$$

where $n = 7$, $r^2 = 0.76$, $s = 0.23$, and $F = 16$. This equation was used to make

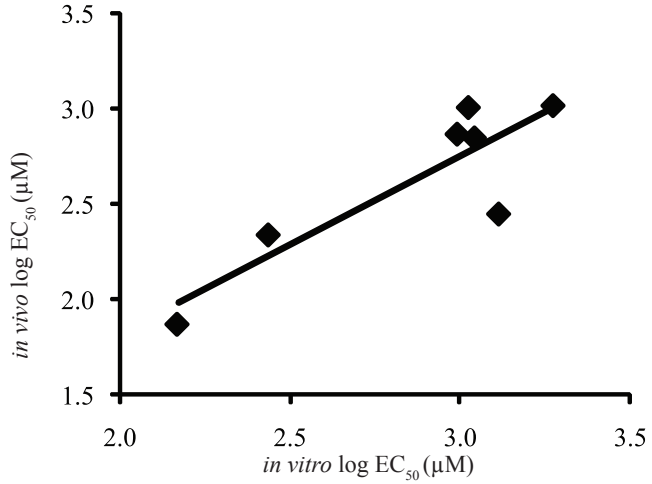


Fig. 3.3: Quantitative structure-activity relationship for the correlation between the decadic logarithm of the 24 h *in vitro* toxicity of chlorinated alkanes to Chinese hamster ovary cells in the MTT assay and 7 d *in vivo* toxicity to fish (*Poecilia reticulata*).

predictions for the compounds tested *in vitro* for which no acute *in vivo* fish toxicity data were available, and the results thus obtained are presented in Tables 3.1 and 3.2.

DISCUSSION

During the development of QSAR models, it is usually recommended where possible to (i) use experimental data from the same laboratory to avoid inter-laboratory variation (*20*) and (ii) use data sets where the ratio of number of test compounds to descriptors used for modeling is at least five (*21*). Both conditions have been satisfied in this study by generating toxicity data within the same laboratory for 26 compounds and developing a QSAR model based on one descriptor, log K_{ow}. The developed QSAR model also satisfies the five basic requirements for OECD guidelines for QSAR models: clearly defined end point, unambiguous algorithm, appropriate measures of goodness of fit, robustness and predictivity, a defined domain of applicability, and a mechanistic interpretation. First, the end points are clearly defined (24 h EC_{50} to CHO cells and 7 d LC_{50} to *P. reticulata*). Second, the methods for data collection and calculation of descriptors have been clearly described. Third, the QSAR model has been validated both internally and externally. Fourth, the applicability domain in terms of descriptor range and the actual list of compounds that fit into the domain have been provided. Fifth, hydrophobicity has been confirmed as an important parameter to describe the toxicity of the chlorinated alkanes. The strong correlation between toxicity and hydrophobicity found in the current study (Figure 3.2) supports a non-polar narcotic mechanism of action for chlorinated alkanes described previously (*12, 22, 23*). Because hydrophobicity is important for the toxicity of chlorinated alkanes, it was essential to choose a suitable method of exposure. The direct method of exposure resulted in higher toxicity than the premix method (Figure 3.1). Hydrophobic compounds have been shown previously to readily adhere to plastic surfaces of culture plates (*24*). This situation can easily arise during the dilution step of the premix method where the medium containing the chlorinated alkanes is prepared in a premixing well before transfer to wells containing the cells. This provides an additional possibility for the chlorinated alkanes to adhere to the plastic surface of the well before the solution is actually transferred to the cells. During the direct exposure method, more of the compound is immediately available to the cells. However, one needs to mix the test compound thoroughly into the culture medium by carefully pipetting up and down several times to avoid any local cytotoxicity that can occur due to high concentrations of solvent or test compound. The presence or absence of local cytotoxicity should always be confirmed with microscopic observations.

Table 3.2: List of EINECS compounds that fit into the applicability domain[a] of the QSAR model developed in this thesis. This list excludes the first 18 compounds in Table 3.1 that were used for the model development. The first five compounds were used as the external validation set.

	Compound name	CAS[b] no.	log K_{ow}[c]	*in vitro* log EC_{50} (μM) to CHO[d] cells		*in vivo* log LC_{50} (μM) to fish *(Poecilia reticulata)*[e]	
				Exp.	Pred[e]	Exp[f]	Pred[g]
1	1,2,3-trichloropropane	96-18-4	1.98	3.12	2.8	2.45	2.87
2	1-chlorobutane	109-69-3	2.52	3.27	2.56	3.02	3.01
3	1,5-dichloropentane	628-76-2	2.77	2.82	2.45		2.59
4	Pentachloroethane	76-01-7	3.63	2.17	2.07	1.87	1.99
5	1,1,1,2,2,2,2,3-heptachloropropane	594-89-8	4.74	1.62	1.57		1.48
6	Hexachloroethane	67-72-1	4.61		1.63		1.49
7	1,1-dichloroethane	75-34-3	1.78		2.89		2.66
8	1,3-dichloro-2,2-bis(chloromethyl) propane	3228-99-7	1.93		2.82		2.59
9	Trichloromethane	67-66-3	1.95		2.81		2.58
10	1-chloropropane	540-54-5	1.99		2.8		2.57
11	2-chloropropane	75-29-6	1.99		2.8		2.57
12	1,3-dichlorobutane	1190-22-3	2.24		2.69		2.47
13	1,4-dichlorobutane	110-56-5	2.24		2.69		2.47
14	1,1-dichloropropane	78-99-9	2.31		2.66		2.44
15	2,2-dichloropropane	594-20-7	2.31		2.66		2.44
16	1,2,3-trichloro-2-methylpropane	1871-58-5	2.38		2.62		2.41
17	1,2-dichloro-2-methylpropane	594-37-6	2.39		2.62		2.4
18	1-chloro-2-methylpropane	513-36-0	2.39		2.62		2.4
19	2-chloro-2-methylpropane	507-20-0	2.39		2.62		2.4
20	1,1,1-trichloroethane	71-55-6	2.48		2.58		2.36
21	1,2,3,4-tetrachlorobutane	3405-32-1	2.5		2.57		2.36
22	2,3-dichlorobutane	7581-97-7	2.52		2.56		2.35
23	2-chlorobutane	78-86-4	2.52		2.56		2.35
24	1,1,3-trichlorobutane	13279-87-3	2.55		2.55		2.34
25	1,1,2,3-tetrachloropropane	18495-30-2	2.57		2.54		2.33
26	1,2,2,3-tetrachloropropane	13116-53-5	2.57		2.54		2.33
27	1,1,2-trichloropropane	598-77-6	2.58		2.53		2.32
28	1,2,2-trichloropropane	3175-23-3	2.58		2.53		2.32
29	1,3-dichloro-3-methylbutane	624-96-4	2.63		2.51		2.3
30	Tetrachloroethane	25322-20-7	2.64		2.51		2.3

	Compound name	CAS[b] no.	log K_{ow} [c]	*in vitro* log EC_{50} (μM) to CHO[d] cells		*in vivo* log LC_{50} (μM) to fish *(Poecilia reticulata)*[e]	
				Exp.	Pred[e]	Exp[f]	Pred[g]
31	1,1,1,3-tetrachloropropane	1070-78-6	2.72		2.47		2.27
32	1,1-dichlorobutane	541-33-3	2.84		2.42		2.22
33	2,2-dichlorobutane	4279-22-5	2.84		2.42		2.22
34	1,1,3,3-tetrachlorobutane	39185-82-5	2.86		2.41		2.21
35	1-chloro-3-methylbutane	107-84-6	2.92		2.38		2.18
36	2-chloro-2-methylbutane	594-36-5	2.92		2.38		2.18
37	Trichloropropane	25735-29-9	3.01		2.34		2.15
38	1,1,1,2-tetrachloroethane	630-20-6	3.03		2.33		2.14
39	2,3-dichloropentane	600-11-3	3.05		2.33		2.13
40	2-chloropentane	625-29-6	3.05		2.32		2.13
41	3-chloropentane	616-20-6	3.05		2.32		2.13
42	1,1,2,2,3-pentachloropropane	16714-68-4	3.17		2.27		2.08
43	1,1,1,3-tetrachlorobutane	13275-19-9	3.25		2.24		2.05
44	1-chloro-3,3-dimethylbutane	8/5/2855	3.32		2.2		2.02
45	1,1,1-trichlorobutane	13279-85-1	3.54		2.11		1.93
46	1,1,1,2-tetrachloropropane	812-03-3	3.56		2.1		1.92
47	1,1,1,3,3-pentachlorobutane	21981-33-9	3.57		2.1		1.92
48	2-chlorohexane	638-28-8	3.58		2.09		1.91
49	3-chlorohexane	2346-81-8	3.58		2.09		1.91
50	2-chloro-2,3,3-trimethylbutane	918-07-0	3.72		2.03		1.85
51	2,5-dichloro-2,5-dimethylhexane	6223-78-5	4.09		1.86		1.7
52	2-chloroheptane	1001-89-4	4.11		1.85		1.69
53	3-chloroheptane	999-52-0	4.11		1.85		1.69
54	1,1,1,3-tetrachloro-4-methylpentane	62103-09-7	4.18		1.82		1.66
55	1-chloro-2,2,4-trimethylpentane	6/4/2371	4.25		1.79		1.64
56	3-(chloromethyl)heptane	123-04-6	4.51		1.67		1.53
57	2-chlorooctane	628-61-5	4.64		1.62		1.47
58	3-chlorooctane	1117-79-9	4.64		1.62		1.47
59	4-chlorooctane	999-07-5	4.64		1.62		1.47

[a] chlorinated alkanes with up to ten carbon atoms and log K_{ow} values between 1.71 and 5.70, [b] Chemical Abstracts Service, [c] octanol-water partition coefficient (K_{ow}) values calculated using ClogP version 4.0 (Biobyte, Claremont, CA, USA), [d] Chinese Hamster Ovary, [e] EC_{50} values predicted using Eq. 2, [f] EC_{50} values obtained from (*12*), [g] EC_{50} values predicted using Eq. 3, [h] Compounds selected for external validation of Eq. 2.

The *in vitro* toxicity of compounds with log K_{ow} > 4.53 (Figure 3.2) can be explained in three ways. First, the predicted solubility values of the chlorinated alkanes are for water only and the solvent used in this study, DMSO, is known to increase their solubility. Second, DMSO may increase the absorption of compounds across membranes (*24, 25*); thus, its use as a co-solvent could enhance their entry into the cells. Third, the test compounds can bind to the proteins or lipids in the FCS in the growth medium, thus increasing their solubility. Previous attempts to use hydrophobicity to explain the toxicity of chlorinated alkanes to bacteria in the Microtox test failed (r^2 = 0.19, n = 18) (*26*), possibly due to the short exposure time of the assay premix exposure as in the present study (Figure 3.1), and also the absence of serum in the medium that could increase the bioavailability of the test compounds. Toxicity results for three of our training set compounds have been reported in other published studies and are comparable to our data. For example, there is close agreement between the cytotoxicity of 1,1,2-trichloroethane, 1,1,2,2-tetrachloroethane, and carbon tetrachloride to CHO cells (current study) and to rat primary hepatocytes, both measured with the MTT assay (Table 3.1). The lower EC_{50} value of carbon tetrachloride to HeLa cells than to the CHO cells of the present study could be due to a longer exposure period (72 h as compared to 24 h in the present study). There is also close similarity between the EC_{50} values obtained in the current study to the concentrations that were toxic to the guppy and fathead minnow (Table 3.1). This similarity was extended to develop a prediction model for *in vivo* toxicity based on *in vitro* or *in silico* data. The good correlation (Figure 3.2, r^2 = 0.88) between the *in silico*-predicted log K_{ow} and the *in vitro* toxicity and the good correlation (Figure 3.3; r^2 = 0.76) obtained between *in vitro* and *in vivo* toxicity to fish support the possible use of QSAR approaches in the safety assessments within the framework of REACH, thereby reducing the use of experimental animals. Therefore, the results of the present study demonstrate that instead of performing toxicity testing of chlorinated alkanes (that fit into the applicability domain) on fish, one can carry out an *in vitro* CHO MTT test or even only calculate the log K_{ow} by available *in silico* models and use the QSAR models defined in the present study. On the basis of the QSAR models that we developed, one can use *in vitro* or even only *in silico* results to predict the *in vivo* toxicity to fish. The experimental and predicted (Tables 3.1 and 3.2) *in vivo* toxicity data to fish can be used as a starting point for further risk assessment of the chlorinated alkanes. A toxicity ranking of the compounds will allow the identification of the most toxic and priority compounds. This will help to direct priorities for future testing to the most toxic compounds, thereby further refining and reducing the use of experimental animals.

ACKNOWLEDGEMENT

We are grateful to The Netherlands Organization for Health Research and Development (ZonMw; Project 3170.0066) for funding this study under the theme "Alternatives to Animal Experimentation".

REFERENCES

(1) EU. (2006) Regulation (EC) No 1907/2006 of the European Parliament and of the Council. *Off. J. Eur. Union*, 1–849.

(2) OECD. (2007) Guidance document on the validation of (Quantitative) structure-activity relationships [(Q)SAR] models, ENV/JM/MONO(2007)2, In *OECD Environment Health and Safety Publications: Series on Testing and Assessment No. 69. Paris, France* pp 1–154.

(3) NRC. (1993) *In Situ* Bioremediation: When Does it Work?, p 33, The National Academy Press, Washington D.C.

(4) Fisk, A. T., Tomy, G. T. and Muir, D. C. G. (1999) Toxicity of C10-, C11-, C12-, and C14-Polychlorinated alkanes to Japanese Medaka (*Oryzias latipes*) embryos. *Environ. Toxicol. Chem. 18*, 2894 - 2902.

(5) Trohalaki, S., Gifford, E. and Pachter, R. (2000) Improved QSARs for predictive toxicology of halogenated hydrocarbons. *Comput. Chem. 24*, 421-427.

(6) UNEP. (2006) Stockholm Convention on Persistent Organic Pollutants review committee: Consideration of chemicals newly proposed for inclusion in annexes A, B, or C of the convention: Short-chained chlorinated paraffins.

(7) Fisk, A. T., Cymbalisty, C. D., Tomy, G. T. and Muir, D. C. G. (1998) Dietary accumulation and depuration of individual C10-, C11- and C14-polychlorinated alkanes by juvenile rainbow trout (*Oncorhynchus mykiss)*. *Aquat. Toxicol. 43*, 209-221.

(8) Crebelli, R., Andreoli, C., Carere, A., Conti, L., Crochi, B., Cotta-Ramusino, M. and Benigni, R. (1995) Toxicology of halogenated aliphatic hydrocarbons: structural and molecular determinants for the disturbance of chromosome segregation and the induction of lipid peroxidation. *Chem. Biol. Interact. 98*, 113-129.

(9) Akers, K. S., Sinks, G. D. and Schultz, T. W. (1999) Structure-toxicity relationships for selected halogenated aliphatic chemicals. *Environ. Toxicol. Pharmacol. 7*, 33–39.

(10) Blaha, L., Damborsky, J. and Nemec, M. (1998) QSAR for acute toxicity of saturated and unsaturated halogenated aliphatic compounds. *Chemosphere 36*, 1345-1365.

(11) Russom, C. L., Bradbury, S. P., Broderius, S. J., Hammermeister, D. E. and Drummond, R. A. (1997) Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). *Environ. Toxicol. Chem. 16*, 948–967.

(12) Könemann, H. (1981) Quantitative structure-activity relationships in fish toxicity studies Part 1: Relationship for 50 industrial pollutants. *Toxicology 19*, 209–221.

(13) Eriksson, L., Jonsson, J. and Berglind, R. (1993) External validation of a QSAR for the acute toxicity of halogenated aliphatic hydrocarbons. *Environ. Toxicol. Chem. 12*, 1185.

(14) Mosmann, T. (1983) Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods. 65*, 55.

(15) Leo, A. (1993) Calculating log Poct from structures. *Chem. Rev. 93*, 1281-1306.

(16) Zvinavashe, E., Murk, A. J., Vervoort, J., Soffers, A. E. M. F., Freidig, A. and Rietjens, I. M. C. M. (2006) Quantum chemistry based quantitative structure-activity relationships for modeling the (sub)acute toxicity of substituted mononitrobenzenes in aquatic systems. *Environ. Toxicol. Chem. 25*, 2313–2321.

(17) Gramatica, P., Pilutti, P. and Papa, E. (2004) Validated QSAR prediction of OH tropospheric degradation of VOCs: splitting into training-test sets and consensus modeling. *J Chem Inf Comput Sci 44*, 1794-1802.

(18) Trohalaki, S., Pachter, R., Geiss, K. T. and Frazier, J. M. (2004) Halogenated aliphatic toxicity QSARs employing metabolite descriptors. *J. Chem. Inf. Comput. Sci. 44*, 1186.

(19) Jaworska, J., Nikolova-Jeliazkova, N. and Aldenberg, T. (2005) QSAR applicability domain estimation by projection of the training set in descriptor space: A review. *ATLA Alternatives*

to Laboratory Animals 33, 445.

(20) Cronin, M. T. D. and Schultz, T. W. (2003) Pitfalls in QSAR. *Journal of Molecular Structure: THEOCHEM 622*, 39-51.

(21) Tropsha, A., Gramatica, P. and Gombar, V. K. (2003) The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR and Combinatorial Science 22*, 69–77.

(22) Sixt, S., Altschuh, J. and Bruggemann, R. (1995) Quantitative structure-toxicity relationships for 80 chlorinated compounds using quantum chemical descriptors. *Chemosphere 30*, 2397-2414.

(23) Veith, G. D., Call, D. J. and Brooke, L. T. (1983) Structure - toxicity relationships for the fathead minnow, *Pimephales promelas*: narcotic industrial chemicals. *Can. J. Fish. Aquat. Sci. 40*, 743–748.

(24) Unger, J. K., Kuehlein, G., Schroers, A., Gerlach, J. C. and Rossaint, R. (2001) Adsorption of xenobiotics to plastic tubing incorporated into dynamic in vitro systems used in pharmacological research - Limits and progress. *Biomaterials 22*, 2031.

(25) Balakin, K. V., Savchuk, N. P. and Tetko, I. V. (2006) *In silico* approaches to prediction of aqueous and DMSO solubility of drug-like compounds: Trends, problems and solutions. *Current Medicinal Chemistry 13*, 223-241.

(26) Blum, D. J. W. and Speece, R. E. (1991) A database of chemical toxicity to environmental bacteria and its use in interspecies comparisons and correlations. *Research Journal of the Water Pollution Control Federation 63*, 198.

Quantitative structure-activity relationship modeling of the toxicity of organothiophosphate pesticides to *Daphnia magna* and *Cyprinus carpio*

Elton Zvinavashe, Tingting Du, Tamas Griff,
Hans H.J. van den Berg, Ans E.M.F. Soffers, Jacques Vervoort,
Albertinka, J. Murk, and Ivonne M.C.M. Rietjens

Submitted

Abstract

Within the REACH regulatory framework in the EU, quantitative structure-activity relationships (QSAR) models are expected to help reduce the number of animals used for experimental testing. The objective of this study was to develop QSAR models to describe the acute toxicity of organothiophosphate pesticides to aquatic organisms. Literature data sets for acute toxicity data of organothiophosphates to fish and one data set from experiments with 15 organothiophosphates on *Daphnia magna* performed in the present study were used to establish quantum chemistry calculation based QSARs. The logarithm of the octanol/water partition coefficient, log $K_{ow,}$ the energy of the lowest unoccupied molecular orbital, E_{lumo}, and the energy of the highest occupied molecular orbital, E_{homo} were used as descriptors. Additionally, it was investigated if toxicity data for the invertebrate *D. magna* could be used to build a QSAR model to predict toxicity to fish. Suitable QSAR models ($0.80 < r^2 < 0.82$) were derived to predict acute toxicity of organothiophosphates to fish (*Cyprinus carpio*) and the crustacean (*D. magna)*. Toxicity data for *D. magna* correlated well ($r^2 = 0.94$) with toxicity data for *C. carpio*. This implies that by performing toxicity tests with *D. magna*, one can use our interspecies QSAR model to predict the acute toxicity of organothiophosphates to fish. The three QSAR models were validated either both internally and externally (*D. magna*) or internally only (carp and *D. magna* to carp). For each QSAR model, an applicability domain was defined based on the chemical structures and the ranges of the descriptor values of the training set compounds. From the 100 196 European Inventory of Existing Commercial Chemical Substances (EINECS), 83 compounds were identified that fit the selection criteria for the QSAR models. For these compounds, using our QSAR models, one can obtain an indication of their toxicity without the need for additional experimental testing.

Introduction

Since June 2007, the management of chemicals within the European Union (EU) falls under a new legislative framework called Registration, Evaluation, and Authorization of Chemicals (REACH) (*1*). REACH aims to provide toxicity information for about 30 000 out of the more than 100 000 chemicals listed on the European Inventory of Existing Commercial Chemical Substances (EINECS), for which there is insufficient toxicological information on their hazardous properties. Within REACH, there are provisions to use, among others, sufficiently validated computational prediction models based on quantitative structure-activity relationships (QSAR) to fill in the toxicity data gaps, and thus save time, money and help to reduce the numbers of animals used for experimental testing purposes. Guidelines for QSAR model development and validation proposed by the Organization for Economic Cooperation and Development (OECD) are expected to help increase the acceptability of QSAR models for regulatory purposes (*2*). This article focuses on the development of validated QSAR models to predict the acute toxicity of organothiophosphate pesticides to fish using toxicity data for invertebrates.

Within the EU, the use of pesticides for pest control in agriculture and animal breeding has greatly increased crop and animal yields and promoted the development of the chemical industry (*3*). Since the ban on organochlorines in the 1970s (*4*), organophosphorous compounds are the most widely used insecticides due to their relatively lower environmental persistence. Because of their widespread use they are often implicated in wildlife and human poisonings (*4*). The general chemical structure of organophosphorous pesticides is shown in Figure 4.1. They are classified into two main groups, organophosphates (P=O) and organothiophosphates (P=S) depending on whether oxygen or sulphur forms a double bond with the central phosphorous atom. Organophosphorous compounds exert their primary acute toxic effects by inhibiting the enzyme acetylcholinesterase (AChE). The inhibition occurs due to a nucleophilic reaction between the hydroxyl group of serine in the active site of AChE and the electrophilic phosphorous atom of the organophosphorous compound resulting in the formation of a covalent P-O bond. As a result AChE can no longer hydrolyze its natural substrate, the neurotransmitter acetylcholine, and acetylcholine accumulates at the synaptic terminals leading to overstimulation of the postsynaptic receptor (*5, 6*).

Since oxygen is more electronegative than sulphur, the P=O bond is more polarized than the P=S bond. This makes the phosphorous atom more electrophilic in the P=O than in the P=S bond (*8*). In the body the P=S moieties of organothiophosphates are oxidized by cytochrome P_{450} monooxygenases to their corresponding P=O or oxon form (*9*), which is the active form

of the compound (*10*). The P=S are manufactured in larger numbers than the P=O as they are considered to be safer and more selective due to the biotransformation step to the P=O that is necessary to exhibit full toxicity (*9*).

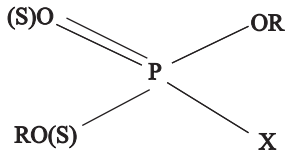


Figure 4.1: General chemical structure of organophosphorous pesticides. The atom that forms a double bond to P is either O (organophosphate; P=O) or S (organothiophosphate; P=S). R is usually methyl or ethyl and X, called the leaving group, can be alkyl, heterocyclic, aryl etc. Adapted from literature (*7*).

Over the last three decades, attempts have been made to develop QSAR models to predict the acute toxicity of organophosphorous compounds to various organisms. Descriptors that have been used include: partitioning parameters, modeled using the logarithm of the octanol-water partitioning coefficient (log K_{ow}) (*11, 12*), steric (*13*) and reactivity parameters (*12-17*). Most of the QSAR models developed in these studies lack a clear definition of their applicability domain and are not externally validated, two of the five basic requirements defined by the OECD guidelines for adequate QSAR development (*2*).

The use of toxicity data for one species to predict toxicity to another species is a promising field that has received little attention. The possibilities for these extrapolations have been demonstrated previously between algae and protozoans (*18*) and between protozoans and fish (*19, 20*). In the present study, the possibilities to use acute toxicity data for *Daphnia magna* (inverterbrate) from our own laboratory experiments to predict acute toxicity of organothiophosphates to fish were investigated. To this end, QSAR models for predicting acute toxicity of organothiophosphates to (i) *D. magna* based on our experimental data and (ii) various fish species based on literature data were developed based on three computer calculated descriptors: log K_{ow}, the energy of the lowest unoccupied molecular orbital (E_{lumo}) and the energy of the highest occupied molecular orbital (E_{homo}). The log K_{ow} generally models a compound's hydrophobicity, which is important in describing the uptake and passage of a compound through membranes. The E_{homo} models the nucleophilic nature of the organothiophosphates, which is important for their oxidation to the P=O form by the electrophilic iron-oxo intermediate of the activated cytochrome P_{450} enzymes. The importance of E_{homo} in modeling cytochrome P_{450} catalyzed reactions has been shown previously (*21, 22*). The E_{lumo} of the P=O metabolite of the various organothiophosphates models the electrophilic

nature of this P=O metabolite, which is important for their covalent interaction with the AChE enzyme. The applicability domain and limits of the QSARs were identified, by noting descriptor value ranges where one can apply the QSARs. The QSAR models were validated both internally and externally. Finally, for the developed QSAR models of the present study, an estimate was made of the number of EINECS compounds for which the models can make accurate predictions**.**

MATERIALS AND METHODS

Acute toxicity tests with D. magna

The acute toxicity of organothiophosphates to *D. magna* was determined in a GLP laboratory following the procedures described in the OECD acute immobilization test (*23*). The test compounds, designated as the training set, were obtained from Sigma-Aldrich (Zwijndrecht, The Netherlands) and were all at least 98 % pure. In order to increase diversity and equal representativity in the chemical structures, the training set compounds were divided into four groups based on (i) the characteristics of the leaving group (see Figure 4.1): aromatic or aliphatic, (ii) components of the reaction centre: OOOP=S or OOSP=S, and (iii) the R-group bound to the reaction centre: ethyl or methyl. Stock solutions of the compounds were prepared in spectrophotometric grade dimethyl sulfoxide (DMSO) obtained from Acros Organics (Geel, Belgium) at concentrations 10 000 times greater than the desired working concentrations in the experiment. Before the start of each experiment, 50 µl of each of the concentrated stock solutions was added to 500 ml of experimental culture medium (*24*) to get the working concentrations. For each compound, six concentrations were tested, each in fourfold. For each concentration, five daphnids, < 24 h old, were placed in each of four vessels with 80 ml of the working solutions. Three sets of controls were used, (i) potassium dichromate as a positive control, (ii) a solvent control with DMSO in place of the stock solution with the test compound, and (iii) culture medium only. O_2 concentration, pH, and temperature were monitored during the experiment and the average values were as follows: O_2 9 mg/ml, pH 7.6, and temperature 20 ^{0}C. The test was considered acceptable when in the solvent and culture medium only controls, not more than 10% of the daphnids should have been immobilized or trapped at the surface of the water, and these conditions were met in our experiments. The number of immobilized daphnids was counted at the end of 24 hours exposure. Immobility was defined as the inability to swim within 15 seconds after gentle agitation of the test vessel (*23*). The nominal concentration of test substance required

to immobilize 50 % of the daphnids after 24 h of exposure (EC_{50}) was estimated using a Microsoft® Excel plug-in, Life Sciences Workbench (LSW) Data Analysis Toolbox Version 1.1.1 (MDL Information Systems, CA, USA) with the general sigmoidal curve with Hill slope (*a* to *d*) chosen as the best fit model.

Literature toxicity data sets

Fifteen experimental toxicity data sets from the literature describing acute toxicity of organothiophosphates to various fish species were screened for their suitability for use as data sets for modeling. The data sets describe acute toxicity of organothiophosphates to six fish species from five genera. For some of the fish species, data from more than one exposure period have been given, e.g. 24 h and 96 h. The fish species for which the acute toxicity data were obtained were rainbow trout *Oncorhynchus mykiss*, bluegill *Lepomis macrochirus* (*25-27*), fathead minnow *Pimephales promelas* (*25, 27*), cutthroat trout *O. clarki*, lake trout *Salvelinus namaycush* (*25*), and carp *Cyprinus carpio* (*7*). Five quality criteria were set in order to guide the selection of suitable data sets: (i) the minimum number of compounds, *n*, should be five ($n \geq 5$) (*28*), (ii) the toxicity data should be from the same laboratory to avoid inter-laboratory variation (*29*), (iii) the toxicity data should have been generated using standardized protocols, if not, then the materials and methods should be sufficiently described in order to judge the quality of the experimental setup, and (iv) the compounds should be easy to identify based on their names and/or chemical abstract (CAS) numbers. Applying these quality criteria resulted in the elimination of 10 data sets from Mayer and Ellersieck (1986).

Calculation of molecular descriptors

Hydrophobicity of the organothiophosphates was modeled using $\log K_{ow}$ values calculated with the software CLogP version 4.0 (Biobyte, Claremont, CA, USA) (*30*) as described previously (*31*). Briefly, the structure of each molecule was entered into CLogP as a simplified molecular input line entry system (SMILES) code (*32*). The SMILES codes were obtained from the SMILES-CAS database (Syracuse Research, Syracuse, NY, USA). Nucleophilicity of the organothiophosphates was modeled using E_{homo} values for the P=S version of each molecule. Electrophilicity of the P=O metabolites of the organothiophosphates was modeled using E_{lumo} values for the P=O version of each molecule. The E_{lumo} and E_{homo} values were calculated using the program Spartan '04 for Linux Server Edition (Wavefun, Irvine, CA, USA; http://www.wavefun.com) running on two 1.7 GHz Xeon Intel CPU's with 4 GB memory. Each molecule was energy minimized, and then a conformer distribution search

was carried out using the Merck Molecular Force Field (MMFF) molecular mechanistic model after which the lowest energy conformer was identified. The E_{lumo} and E_{homo} values were calculated for the lowest energy conformer without further geometry optimization using the *ab initio* Hartree-Fock method utilizing a 3-21G basis set.

QSAR modeling

The Statistical Package for Social Scientists (SPSS®) version 15 for Windows (SPSS, Chicago, IL, USA) was used to analyze the QSAR models as described previously (*31*). Briefly, the log (toxicity endpoint) was the dependent variable while log K_{ow}, E_{homo}, and E_{lumo} were used singly or in combination as the independent variables. All QSAR equations were presented in the form:

$$\text{predicted log } EC_{50}\text{ (M)} = x(\log K_{ow}) + y(E_{lumo}) + z(E_{homo}) + c \qquad [1]$$

where *x, y*, *z* and *c* are fitting parameters. The quality of the QSAR models was characterized by the number of compounds used in the study (*n*), coefficient of determination (r^2), standard error of the estimate (*s*), variance ratio (*F*), and the internally cross-validated coefficient of determination (r^2_{int}). Where a test set was available, the externally cross-validated coefficient of determination (r^2_{ext}) was also determined. Internal validation of the QSAR models was performed using the leave-one-out (LOO) validation method where $n \leq 10$. The leave-many-out (LMO) validation method, was used when $n > 10$, with 20 % of the calibration compounds left out at each step as described previously (*31*). Compounds whose studentized deleted residuals were outside the $\pm$ 1.5 limit were denoted as outliers (*33*) during the modeling process. Internal validation was only performed for QSAR models where $r^2 \geq 0.65$ and $n/k \geq 5$, where *k* is the number of descriptors in the model (*28*). The chemical applicability domain of the QSAR models was defined by taking into account two aspects of the compounds in the training set (i) the range of values for the descriptors used for modeling and (ii) the presence of P=S in the chemical structure.

External validation of QSAR model for acute toxicity to D. magna.

The external performance of the *D. magna* QSAR model based on 10 compounds was evaluated by testing five compounds that cover the applicability domain of the model, and then comparing the predicted and obtained experimental toxicity values. Selection of test concentrations was based on EC_{50} values predicted by the QSAR model developed using the training set data. The calculation of r^2_{ext} was performed as described previously (*34*).

RESULTS

D. magna acute toxicity test

The results of fifteen organothiophosphates tested in the *D. magna* acute immobilization test are presented in Table 4.1 (compounds 1 to 15), and include the experimental EC_{50} values obtained together with their estimated log $K_{ow,}$ $E_{homo,}$ the E_{lumo} of their P=O metabolites. The first 10 compounds in Table 4.1 were used as the training set for building a QSAR model and their log EC_{50} (M) values ranged almost five orders of magnitude (from -5.06 to -9.82). A higher logarithmic toxicity value (e.g. -5.06) represents lower toxicity than a lower logarithmic toxicity value (e.g. -9.82).

QSAR modeling: D. magna toxicity data set

Figure 4.2 presents a plot of the decadic logarithm of the experimental log EC_{50} values for the 10 organothiophosphates in the training set against values predicted based on the estimated log K_{ow}.

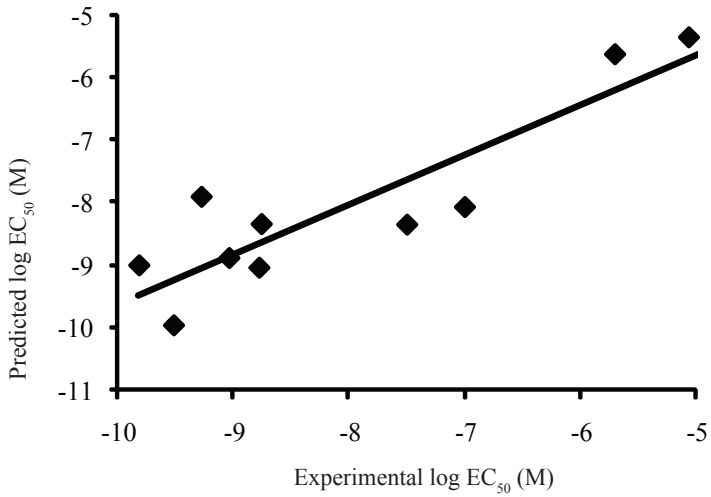


Figure 4.2: Quantitative structure-activity relationship for the correlation between the decadic logarithm of the 24 h *in vitro* toxicity (EC_{50}) of organothiophosphates to *Daphnia magna* and the toxicity predicted based on log K_{ow} (n = 10, r^2 = 0.80, r^2_{int} = 0.73)

Table 4.1: Chemical Abstract Service (CAS) numbers, chemical classification scheme, calculated descriptors, and experimental toxicity data for organothiophosphates to *Daphnia magna and Cyprinus carpio* (carp)

No[a].	CAS no.	Compound	Compound classification[b]			log K_{ow}[c]	E_{homo}[d] $_{(P=S)}$[e]	E_{lumo}[f] $_{(P=O)}$[g]	*D. magna* 24 h log EC_{50} (M)		Carp Exp. 14 d
			R group	leaving group	reaction center		in eV	in eV	Exp.[h]	Predicted[i]	log LC_{50} (M)[j]
1	2540-82-1	Formothion	methyl	aliphatic	OOSP=S	1.03	-9.61	2.23	-5.7	-5.61	
2	14816-18-3	Phoxim	ethyl	aromatic	OOOP=S	4.39	-9.49	1.19	-9.04	-8.88	
3	2463-84-5	Dicapthon	methyl	aromatic	OOOP=S	3.55	-9.99	0.9	-7	-8.06	
4	5598-13-0	Chlorpyriphos methyl	methyl	aromatic	OOOP=S	3.83	-9.62	1.65	-8.76	-8.34	-5.19
5	55-38-9	Fenthion	methyl	aromatic	OOOP=S	3.84	-8.92	3.12	-7.5	-8.35	-5.14
6	2921-88-2	Chlorpyriphos ethyl	ethyl	aromatic	OOOP=S	4.51	-9.58	1.6	-9.82	-9	
7	29232-93-7	Pirimiphos methyl	methyl	aromatic	OOOP=S	3.38	-8.55	3.3	-9.28	-7.9	
8	18181-70-9	Iodofenphos	methyl	aromatic	OOOP=S	5.5	-9.1	2.54	-9.52	-9.96	
9	60-51-5	Dimethoate	methyl	aliphatic	OOSP=S	0.75	-9.45	3.92	-5.06	-5.34	
10	13071-79-9	Terbufos	ethyl	aliphatic	OOSP=S	4.55	-9.19	4.24	-8.78	-9.04	
11	121-75-5	Malathion	methyl	aliphatic	OOSP=S	2.48	-9.46	3.72	-7.3	-7.02	-4.57
12	298-02-2	Phorate	ethyl	aliphatic	OOSP=S	3.84	-9.26	4.15	-8.17	-8.35	
13	333-41-5	Diazinon	ethyl	aromatic	OOOP=S	3.53	-9.23	2.98	-8.3	-8.04	-4.98
14	122-14-5	Fenitrothion	methyl	aromatic	OOOP=S	3.21	-9.76	1.64	-8	-7.73	-4.83
15	298-00-0	Parathion methyl	methyl	aromatic	OOOP=S	2.79	-9.93	1.19	-7.2	-7.32	
16	2636-26-2	Cyanophos	methyl	aromatic	OOOP=S	2.51	-9.53	2.41		-7.04	-4.67
17	3811-49-2	Dioxabenzofos	methyl	aromatic	OOOP=S	2.67	-9.24	3.29		-7.2	-4.79
18	18854-01-8	Isoxathion	ethyl	aromatic	OOOP=S	4.03	-9.03	1.83		-8.52	-5.17
19	950-37-8	Methidathion	methyl	aliphatic	OOSP=S	2.77	-9	3.75		-7.3	-5.08
20	56-38-2	Parathion ethyl	ethyl	aromatic	OOOP=S	3.47	-9.88	1.2		-7.98	
21	2597-03-7	Phenthoate	methyl	aromatic	OOSP=S	3.83	-9.23	2.96		-8.33	-5.21

[a] Compounds 1—10 were the training set compounds for the *D. magna* model, and compounds 11–15 were the validation set. [b] see Figure 1 for information on the position of the R-group, leaving group and reaction center on the molecule. [c] logarithm of the octanol-water partition coefficient. [d] energy of the highest occupied molecular orbital in electron volts [e] organothiophosphate [f] energy of the lowest unoccupied molecular orbital in electron volts [g] organophosphate (oxidized form of the organothiophosphate) [h] values obtained from experiments in the present study, [i] values predicted using Eq. 3 in Table 4.2. Experimental toxicity data from [j] (7).

Table 4.2: Quantitative structure-activity relationships (QSARs) for predicting toxic endpoints for *Daphnia magna* and *Cyprinus carpio* exposed to organothiophosphate pesticides, using the octanol/water partition coefficient (K_{ow}), energy of the lowest unoccupied molecular orbital (E_{lumo}), and the energy of the highest occupied molecular orbital as descriptors (E_{homo})

Equation	Toxicity endpoint	log toxicity (M) = x(log K_{ow}) + y(E_{lumo}) + z(E_{homo}) + c	n[a]	r^2[b]	s[c]	F[d]	r^2_{int}[e]	r^2_{ext}[f]	Ref for tox data
3	24 h EC_{50}[g] *Daphnia magna*	-0.974(log K_{ow})- 4.598	10	0.8	0.79	32	0.73	0.71	this thesis
		(log K_{ow} range: 0.75 to 5.50)							
4	24 h EC_{50} *D. magna*	-0.975(log K_{ow}) - 0.007(E_{lumo}) - 10.583	10	0.8	0.84	16	0.66	0.65	this thesis
		(log K_{ow} range: 0.75 to 5.50)							
		(E_{lumo} range: 0.90 to 4.24)							
5	24 h EC_{50} *D. magna*	-0.938(log K_{ow}) - 0.639(E_{homo}) - 7.294	10	0.82	0.79	14	0.62	0.61	this thesis
		(log K_{ow} range: 0.75 to 5.50)							
		(E_{homo} range -9.99 to -8.92)							
6	14 d LC_{50}[i] *Cyprinus carpio*	-0.367(log K_{ow}) - 0.064(E_{lumo}) – 3.587	10	0.8	0.12	12	0.6		*(7)*
		(log K_{ow} range: 2.48 to 4.03)							
		(E_{lumo} range: 1.64 to 3.75)							

[a] number of compounds used in QSAR model [b] coefficient of determination. [c] standard error of the estimate. [d] variance ratio. [e] cross-validated correlation coefficient. [f] externally validated coefficient of determination. [g] EC_{50} = effective concentration to 50 % of test organisms.
[h] LD_{50} = dose lethal to 50 % of test organisms. [i] LC_{50} = concentration lethal to 50 % of test organisms.

Figure 4.2 shows that all data points are within one log unit of the regression line. Table 4.2 shows the results of the correlations between various descriptors and experimental toxicity to *D. magna* (Eqns 3–5). The log K_{ow} correlated well with experimental toxicity (Eqn. 3; r^2 = 0.80). The addition of E_{lumo} as a second descriptor did not improve the correlation (Eqn. 4; r^2 = 0.80). Combining log K_{ow} with E_{homo} resulted in a slightly better correlation (Eqn. 5; r^2 = 0.82) than with log K_{ow} only (Eqn. 3; r^2 = 0.80). During the internal validation of the three QSAR models for *D. magna* (Table 4.2; Eqns 3–5*)*, the model with only log K_{ow} as a descriptor performed best (Eqn. 3; r^2_{int} = 0.73) when compared to the models with E_{lumo} (Eqn. 4; r^2_{int} = 0.66) or E_{homo} (Eqn. 5, r^2_{int} = 0.62) as additional descriptors to log K_{ow}. Therefore, the QSAR model with log K_{ow} only was selected for external validation experiments.

External validation of D. magna QSAR model

For external validation tests, five compounds were selected to cover the log K_{ow} range of the training set compounds (Table 4.1, compounds 11 to 15). Test concentrations were selected based on toxicity values predicted using the QSAR model with log K_{ow} as descriptor (Table 4.2, Eqn. 3). The experimental toxicity values shown in Table 4.1, correlated well with predicted values (r^2_{ext} = 0.71).

Modeling the fish toxicity data sets

Out of all the fish toxicity data sets, an acceptable QSAR model ($r^2 \geq 0.65$) was only obtained for carp (Table 4.2; Eqn. 6; r^2 = 0.80, r^2_{int} = 0.60) using a combination of both log K_{ow} and E_{lumo} (Eqn. 6; r^2 = 0.80, r^2_{int} = 0.60). In this chapter, we have presented only the experimental toxicity data for carp (Table 4.1), while the rest of the toxicity data sets are available as an appendix at the end of this thesis.

Applicability domain of the QSAR models

The range of EINECS compounds for which the two QSAR models (*D. magna* and carp) are valid (applicability domain) was determined by taking into account (i) the minimum and maximum values of the descriptors in the data sets used during the modeling process and (ii) the presence of P=S in the molecular structure. For the two QSAR models, the minimum and maximum values for the descriptors are shown in Table 4.2 and they were set as follows: *D. magna* (Eqn. 3; log K_{ow} range 0.75 to 5.50) and carp (Eqn. 6; log K_{ow} range 2.48 to 4.03, E_{lumo} range 1.64 to 3.75 eV).

The descriptor ranges and presence of P=S in the molecular structure were combined with the molecular formula, compound name, and SMILES codes in procedures described

previously (*31*) to extract compounds with similar properties from the EINECS list. Table 4.3 shows the extracted compounds together with their estimated log K_{ow}, E_{lumo}, $E_{homo,}$ and EC_{50} values. EC_{50} values are not indicated where the compound does not fit into the applicability domain of the specific QSAR model. For example, no predicted EC_{50} value for phoxim (compound 12) to carp is indicated because its log K_{ow} value of 4.39 means that it lies outside the applicability domain of the carp QSAR model (Table 4.2, Eqn. 6, log K_{ow} range 2.48 to 4.03). The QSAR models cover 68 (*D. magna*) and 14 (carp) EINECS compounds in addition to the ones used as training and test set compounds (Table 4.3). Including the training and test set compounds shown in Table 4.1, each QSAR model covers 83 (*D. magna*) and 24 (carp) EINECS compounds. This implies that including the 21 compounds in Table 4.1 and taking into account overlap between the EINECS compounds covered by each QSAR, the two QSAR models cover 83 compounds, which amounts to ~0.1 % of the total EINECS list.

Interspecies correlations

In a final step, it was investigated whether acute toxicity data for the invertebrate *D. magna* could be used to build a model for making *in vivo* toxicity predictions to the vertebrate carp. Figure 4.3 presents a plot of experimental EC_{50} values for carp versus predicted EC_{50} values based on the QSAR for *D. magna.* From Figure 4.3 it can be seen that methidathion is an outlier, and this is confirmed by its high studentized deleted residual value of 4, which is outside the set limit of $\pm$ 1.5. Repeating the QSAR modeling in the absence of methidathion resulted in an improved correlation (r^2 = 0.94) than in the presence of methidathion (r^2 = 0.74). The plot in Figure 4.3 excludes methidathion and can be described by the following equation:

$$\log EC_{50}\,(M)\ (carp) = 0.39[\log EC_{50}\,(M)\ \textit{D. magna}] - 1.93 \qquad [2]$$

where $n = 9$, $r^2 = 0.94$, $r^2_{int} = 0.90$, $s = 0.06$, $F = 107$

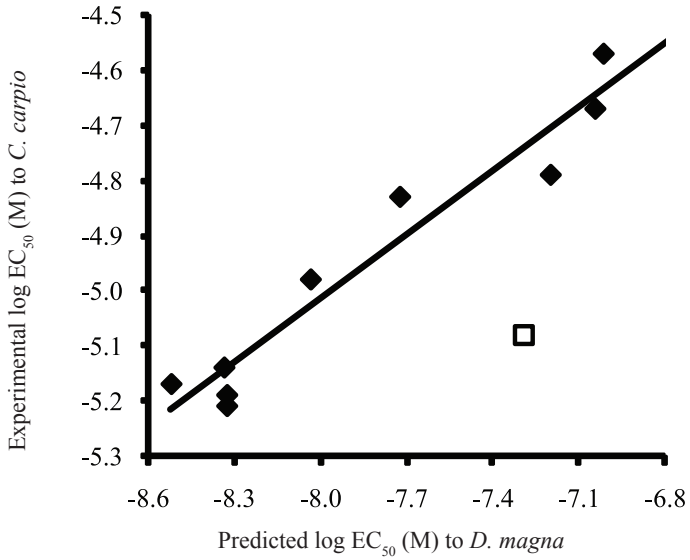


Figure 4.3: Quantitative structure-activity relationship for the correlation between the decadic logarithm of the 14 d experimental LC_{50} values of organothiophosphates to *Cyprinus carpio* (carp) and EC_{50} values predicted using the QSAR for *Daphnia magna* based on log K_{ow} ($n = 9$, $r^2 = 0.94$, $r^2_{int} = 0.90$). Note that the plot only includes the nine compounds (♦) whose studentized deleted residual (SDR) lies within the $\pm$ 1.5 limit. Methidathion (□), whose SDR = 4, is considered an outlier and is not included in the plot.

Table 4.3: List of EINECS compounds that fit into the applicability domain of one or more of the two single species QSAR models developed in the present thesis, their Chemical Abstract Services (CAS) numbers and predicted EC_{50} values.

No	CAS no.	Compound name	log K_{ow} [a]	E_{lumo} [b] (P=O)[e] eV)	E_{homo} [d] (P=S) [e](eV)	log EC_{50} (M)[f] *Daphnia magna* 24 h	log LC_{50} (M)[h] *Cyprinus carpio* 14 d
1	10311-84-9	dialifor	4.36	4.36	-6.71	-8.84	
2	107-55-1	O,O-di-sec-butyl hydrogen dithiophosphate	3.67	4.19	-9.44	-8.17	
3	107-56-2	O,O-diisopropyl hydrogen dithiophosphate	2.61	4.19	-9.46	-7.14	
4	115-90-2	fensulfothion	2.24	3.1	-9.1	-6.78	
5	115-93-5	proban	1.23	2.52	-9.65	-5.8	
6	116-01-8	dimethoate-ethyl	1.28	3.91	-9.25	-5.84	
7	119-12-0	pyridafenthion	2.51	1.6	-8.4	-7.04	
8	126-68-1	ethyl-phosphorothioate	2.02	6.79	-9.51	-6.57	
9	13457-18-6	pyrazophos	3.07	1.96	-8.92	-7.59	-4.84
10	13593-03-8	quinalphos	3.2	1.17	-9.07	-7.71	
11	144-41-2	morphothion	1.3	4.3	-9.31	-5.86	
12	14816-18-3	phoxim	4.38	1.19	-9.49	-8.86[i]	
13	14816-20-7	chlorphoxim	5.1	2.36	-9.25	-9.57	
14	1497-32-1	thiobenzoic acid, anhydrosulphide with O,O-diethyl- dithiophosphorate	4.57	1.64	-9.28	-9.05	
15	1716-09-2	ethyl-fenthion	4.52	3.15	-8.9	-9	
16	18854-01-8	isoxathion	4.03	1.83	-9.03	-8.52	-5.18[i]
17	2104-96-3	bromophos	5.1	2.5	-9.39	-9.57	
18	22068-88-8	neosar	3.11	1.88	-9.39	-7.63	-4.85
19	2253-43-2	O,O-dipropyl hydrogen dithiophosphate	3.05	4.19	-9.46	-7.57	
20	2253-44-3	O,O-dibutyl hydrogen dithiophosphate	4.11	4.23	-9.36	-8.6	
21	2253-52-3	O,O-diisobutyl hydrogen dithiophosphate	3.85	4	-9.48	-8.35	
22	2275-18-5	prothoate	2.47	3.82	-9.19	-7	
23	2310-17-0	phosalone	4.31	2.8	-9.05	-8.8	
24	23505-41-1	pirimiphos-ethyl	4.06	3.31	-8.61	-8.55	
25	24017-47-8	triazophos	2.9	2.39	-8.88	-7.42	-4.8
26	24934-91-6	chlormephos	3.21	3.78	-9.53	-7.72	

No	CAS no.	Compound name	log K_{ow} [a]	E_{lumo} [b] (P=O)[c] eV)	E_{homo} [d] (P=S) [e](eV)	log EC_{50} (M)[f] *Daphnia magna* 24 h	log LC_{50} (M)[h] *Cyprinus carpio* 14 d
27	7/6/2497	disulfoton sulfoxide	2.31	4.13	-9.33	-6.85	
28	2595-54-2	mecarbam	3.26	3.61	-9.28	-7.77	-5.01
29	2597-03 7	phenthoate	3.84	2.96	-9.23	-8.34	
30	2636-26-2	cyanophos	2.5	2.41	-9.53	-7.03	-4.66[i]
31	2642-71-9	azinphos-ethyl	3.43	1.53	-9.29	-7.94	
32	2921-88-2	chlorpyrifos (-ethyl)	4.51	1.6	-9.58	-8.99[i]	
33	29232-93-7	pirimiphos-methyl	3.38	3.3	-8.55	-7.89[i]	-5.04
34	297-97-2	cynophos, thionazin	1.6	2.15	-9.36	-6.16	
35	298-02-2	phorate	3.84	4.15	-9.62	-8.35	
36	298-03-3	demeton-O, demetonthione	2.64	5.65	-9.25	-7.17	
37	298-04-4	disulfoton	4.17	4.37	-9.25	-8.66	
38	298-06-6	O,O-diethyl hydrogen phosphorodithioate	1.99	4.15	-9.5	-6.54	
39	299-84-3	fenchlorphos	4.97	2.46	-9.48	-9.44	
40	3070-15-3	fensulfothion sulfide	4.02	3.42	-8.68	-8.51	-5.28
41	30864-28-9	methacrifos	0.98	3.07	-9.67	-5.55	
42	32345-29-2	ethyl-phenyl-thiophosphate	3.42	3.62	-9	-7.93	-5.07
43	32650-55-8	O,O'-diisopentyl hydrogen dithiophosphate	4.9	4.12	-9.52	-9.37	
44	33712-72-0	tris(2-chloropropyl) thiophosphate	3.8	4.19	-9.76	-8.3	
45	34643-46-4	prothiofos	5.38	2.76	-9.26	-9.84	
46	35400-43-2	sulprofos	5.48	3.27	-8.65	-9.94	
47	36145-08-1	chlorprazophos	3.27	2.25	-8.74	-7.78	-4.93
48	36560-17-5	O-(2,2-dichlorovinyl) O,O-dimethylthiophosphate	1.82	3.76	-9.76	-6.37	
49	3811-49-2	dioxabenzofos	2.67	3.29	-9.24	-7.2	-4.78[i]
50	38260-54-7	etrimphos	3.41	3.08	-9.39	-7.92	-5.04
51	42509-80-8	isazofos	2.51	3.72	-9.39	-7.04	-4.75
52	500-28-7	chlorothion	3.4	1.8	-10.02	-7.91	-4.95
53	5221-49-8	pyrimitate	3	3.2	-8.86	-7.52	-4.69
54	52664-35-4	tris(4-aminophenyl) thiophosphate	2.69	3.73	-7.85	-7.22	-4.81
55	52-85-7	famophos	2.19	2.52	-9.67	-6.73	
56	55-38-9	fenthion	3.84	3.12	-8.92	-8.34[i]	-5.20[i]

No	CAS no.	Compound name	log K_{ow} [a]	E_{lumo} [b] (P=O)[c] eV)	E_{homo} [d] (P=S) [c](eV)	log EC_{50} (M)[f] *Daphnia magna* 24 h	log LC_{50} (M)[h] *Cyprinus carpio* 14 d
57	56-38-2	parathion (-ethyl)	3.47	1.2	-9.88	-7.98	
58	56-72-4	coumaphos	4.33	1.48	-9.02	-8.82	
59	57018-04-9	tolclofos-methyl	4.86	2.88	-9.38	-9.33	
60	6028-46-2	O,O-dicyclohexyl hydrogen dithiophosphate	4.99	4.16	4.16	-9.46	
61	6044-12 8	O,O,O-tris(2-chloro-1-methylethyl) phosphorothioate	3.14	5.29	-10.03	-7.66	
62	640-15-3	thiometon	3.2	4.33	-9.26	-7.71	
63	64249-01-0	anilophos	4.36	3.08	-9.07	-8.84	
64	68715-90-2	O-sec-butyl O-ethyl hydrogen dithiophosphate	2.83	4.14	-9.47	-7.35	
65	68957-49-3	O-[1-(tert-butyl)-5-chloro-1H-1,2,4-triazol-3-yl]-O,O-diethyl thiophosphate	2.91	3.56	-9.37	-7.43	-4.88
66	71735-74-5	ethyl 3-[[bis(1-methylethoxy) phosphinothioyl]thio] propionate	4.13	4.4	-9.35	-8.62	
67	732-11-6	phosmet	3.11	1.26	-9.35	-7.63	
68	756-80-9	O,O-dimethyl hydrogen dithiophosphate	0.93	4.09	-9.55	-5.5	
69	757-86-8	methyl [(dimethoxyphosphinothioyl) thio]acetate	1.78	4.21	-9.5	-6.33	
70	78-34-2	dioxathion	4.01	3.93	-9.46	-8.5	
71	78-57-9	azidithion	0.79	2.83	-10.1	-5.37	
72	86-50-0	azinphos methyl	2.55	1.55	-9.42	-7.08	
73	950-37-8	methidathion	2.77	3.75	-9.38	-7.3	-4.84[i]
74	97-17-6	dichlofenthion	5.04	2.88	-9.83	-9.51	

[a] octanol/water partition coefficient [b] energy of the lowest unoccupied molecular orbital in electron volts [c] organophosphate (oxidized form of the organothiophosphate) [d] energy of the highest occupied molecular orbital in electron volts [e] organothiophosphate cross-validated correlation coefficient. [f] EC_{50} values predicted using [f]Eqn.3 (*D. magna*), [g]Eqn. 6 (honeybee), and [h]Eqn. 7 (carp). EC_{50} values not predicted for compounds that do not fit into the applicability domain (blank EC_{50} values). [i] These compounds already appear in Table 4.1.

DISCUSSION

In the present study, we have investigated the possibilities of developing a validated interspecies QSAR model for predicting the acute toxicity of organothiophosphates to fish using toxicity data for *D. magna*. We have successfully developed an externally validated QSAR model for predicting 24 h acute toxicity to *D. magna* based on toxicity data generated in our laboratory experiments and used it to develop an interspecies QSAR model to make toxicity predictions to carp. The preference to carry out toxicity tests on *D. magna* was based on several considerations. First, within the REACH regulation, *Daphnia* is one of the two species (together with algae) that are preferred species for providing short-term toxicity data (*1*). Second, *Daphnia* was preferred over algae species because (a) algal experiments are known to result in experiments with relatively high variability due to the large diversity of methods and protocols used for toxicity determination (*35*) and (b) algae are unicellular, thus have no nerve cells, which are relevant when one wishes to consider the neurotoxicity of organothiophosphates. Third, for *Daphnia*, there is a standardized OECD toxicity test available (*23*). Fourth, attempts to carry out external validation experiments based on a QSAR model developed on existing *D. magna* literature data (*17*) (data not shown) failed. This could be probably due to variations or differences in the *D. magna* clones used by Vighi et al 17 years ago with the clone we used, that could result in wide deviations in the expected toxicity response. Therefore, it was preferable to generate our own toxicity data for QSAR modeling and for external validation. Based on the set of toxicity data generated a QSAR model according to OECD defined guidelines could be defined.

The OECD recommends that for each QSAR model, the following five categories of information be provided: i) a clearly defined endpoint, ii) an unambiguous algorithm, iii) appropriate measures of goodness of fit, robustness, and predictivity, iv) a defined domain of applicability, and v) a mechanistic interpretation (*2*). Since detailed explanations of the OECD guidelines have been published (*2*) and are freely available, only a brief description of the application of the five criteria in this study will be provided. The QSAR models developed in the present study satisfy all five (*D. magna)* and four (carp) of the OECD requirements. Since a given toxicity endpoint could be determined under different experimental conditions and protocols (*2*), it is important to provide all information about the experimental setup and the measured endpoint, as done in the present study [24 h EC_{50} (M) for *D. magna* and 14 d LC_{50} (M) for carp)]. For each of the three descriptors: log K_{ow}, E_{lumo} and E_{homo}, relevance to toxicity endpoint, calculation methods, software packages and the predicted values of the descriptors for each compound are described and the latter were generated

by methods that are publicly available. The statistical technique, linear regression used to analyze the relationship between toxicity and the physicochemical descriptors is simple to use and sufficient especially where few descriptors (in our case, two) are involved (*36*). This ensures transparency of the model algorithm. The QSAR models have been validated both internally (Eqn. 3; $r^2_{int} = 0.73$; Eqn. 6; $r^2_{int} = 0.60$), a measure of robustness) and externally for *D. magna* (Eqn. 3; $r^2_{ext} = 0.71$, a measure of predictivity). As models are a simplification of reality, their limits should be well defined (*37*). In our study, the applicability domain in terms of both the descriptor range (for the *D. magna* QSAR model) and the actual list of compounds that fit into the domain have been provided (see Table 4.3). Hydrophobicity has been identified as an important parameter to describe the toxicity of P=S compounds to *D. magna*. An additional parameter, E_{homo}, slightly improves the toxicity prediction for *D. magna* (Eqn. 5; $r^2 = 0.82$). When comparing the correlation with log K_{ow} only (Eqn. 3; $r^2 = 0.80$) and that with log K_{ow} + E_{homo} (Eqn. 5; $r^2 = 0.82$) the only marginal improvement in the correlation when taking E_{homo} into account as an additional descriptor, reaffirms the fact that toxicity of organothiophosphates to *D. magna* is limited by the ability of the compounds to enter the organism (modeled by log K_{ow}) as shown previously (*17*). The slight improvement in the correlation after the addition of E_{homo} and the lack of improvement after adding E_{lumo} suggests that for *D. magna*, the conversion of the organothiophosphate to the organophosphate (modeled by E_{homo}) or the rate of reaction of the P=O metabolite of the organothiophosphate with AChE (modeled by E_{lumo}) are not the rate-limiting steps in the mechanism underlying the toxicity. Toxicity to carp is dependent on both the entry of the organothiophosphates into the organism modeled by log K_{ow}, and their conversion to the active oxon (P=O) (modeled by E_{lumo}) form.

The good interspecies correlation ($r^2 = 0.94$) obtained in this study for *D. magna* to carp, seems to support the idea that one can use toxicity data for one organism to predict toxicity to another organism. Our correlation compares well with that from previous studies using protozoan data to predict toxicity to fish for aldehydes ($r^2 = 0.81$) (*19*) and common organic pollutants ($r^2 = 0.81$) (*20*).

The results in the present study can be used in several ways to help reduce the number of animals used for experimental purposes. In case of new organothiophosphates that fit our defined selection criteria, the defined toxicological effects to *D. magna* and carp can be estimated using our developed models without any additional animal testing. The experimental and predicted EC_{50} values (Table 4.1 and 3) can be used to identify the most toxic organothiophosphates on which future risk assessment steps should concentrate. In case further experiments are necessary one can decide to perform experiments with *D. magna*

and then use the *D. magna* to carp QSAR model (Eq. 2) to predict toxicity to carp. Predicted or experimental EC_{50} values for *D. magna* can also be used in the acute threshold test (*38*). This test is based on the experience that daphnids are more sensitive to toxicants than fish; therefore, the EC_{50} values for *D. magna* can be used as an upper threshold for fish tests (*38*). Using the upper threshold concentration, tests with fish can be repeated until a concentration is reached at which no mortality is found – the so-called "step down procedure", which can result in a reduction of approximately 73 % in the number of fish used (*38*). Hoekzema et al recently validated the test and demonstrated its usefulness for various groups of chemicals and came to a reduction of 88 % of the number of fish used (*39*). All these approaches support the reduction of the use of experimental animals, which is one of the goals of REACH (*1*).

QSAR models are expected to play an important role in the risk assessment of chemicals within REACH. To allow a trustable application of QSAR models on the regulatory scene, the implementation of the OECD guidelines for QSAR development and validation at every step of the modeling process is very important. Whenever additional data sets are available, it is preferable that QSAR models are externally validated. The three QSAR models described in the present paper can altogether cover 83 compounds, equivalent to 0.1 % of the EINECS list.

Appendices

Supplementary information is available in three tables (S1-S3) on page 137 .

Acknowledgements

We are grateful to the Netherlands Organization for Health Research and Development, (ZonMw; project no. 3170.0066) for funding this study under the theme "Alternatives to animal experimentation". We also thank NOTOX B.V. NL for (i) providing facilities to perform the *Daphnia* acute toxicity experiments and (ii) for the excellent technical assistance from Erik de Vries, and Vicky Carolus.

References

(1) EU. (2006) Regulation (EC) No 1907/2006 of the European Parliament and of the Council. *Off. J. Eur. Union*, 1–849.

(2) OECD. (2007) Guidance document on the validation of (Quantitative) structure-activity relationships [(Q)SAR] models, ENV/JM/MONO(2007)2, In *OECD Environment Health and Safety Publications: Series on Testing and Assessment No. 69. Paris, France* pp 1–154.

(3) EU. (2001) White Paper: Strategy for a future chemicals policy. Commission of the European Committees, Brussels, Belgium. COM(2001) 88 final., pp 1–32.

(4) EPA. (1999) Recognition and Management of Pesticide Poisonings, (Reigart, J. R. and Roberts, J. R., Eds.) p 243.

(5) Vittozzi, L., Fabrizi, L., Di Consiglio, E. and Testai, E. (2001) Mechanistic aspects of organophosphorothionate toxicity in fish and humans. *Environ. Int. 26*, 125–129.

(6) Nigg, H. N. and Knaak, J. B. (2000) Blood cholinesterases as human biomarkers of organophosphorous pesticide exposure. *Rev. Environ. Contam. Toxicol. 163*, 29–111.

(7) Saito, H., Iwami, S. and Shigeoka, T. (1991) In vitro cytotoxicity of 45 pesticides to goldfish GF-scale (GFS) cells. *Chemosphere 23*, 525-537.

(8) Fukuto, T. R. (1990) Mechanism of action of organophosphorous and carbamate insecticides. *Environ. Health Perspect. 87*, 245–254.

(9) Jokanovic, M. (2001) Biotransformation of organophosphorous compounds. *Toxicology 166*, 139-160.

(10) Drew, G. B., Lumley, J. A. and Price, N. R. (1999) Predicting ecotoxicology of organophosphate insecticides: successful parameter selection with the genetic function algorithm. *Quant. Struct. Act. Relat. 18*, 573–583.

(11) Simon, Z. (1980) Steric fit in quantitative structure-activity relations., In *Lecture notes in Chemistry* (Balaban, A. T., Chiriac, I., Motoc, and Simon, Z., Ed.) pp 67-73, Springer-Verlag, Berlin.

(12) Schuurmann, G. (1990) QSAR analysis of the acute fish toxicity of organic phosphorothionates using theoretically derived molecular descriptors. *Environ. Toxicol. Chem. 9*, 417-428.

(13) Singh, A. K. (1999) Quantitative structure-activity relationships for phosphoramidothioate toxicity in housefly. *Comp. Biochem. Physiol. C: Pharmacol. Toxicol. 123*, 241-255.

(14) Bermudez-Saldana, J. M. and Cronin, M. T. D. (2006) Quantitative structure-activity relationships for the toxicity of organophosphorous and carbamate pesticides to the rainbow trout *Onchorhyncus mykiss*. *Pest Manag. Sci. 62*, 819–831.

(15) De Bruijn, J. and Hermens, J. (1991) Qualitative and quantitative modeling of toxic effects of organophosphorous compounds to fish. *Sci. Total Environ. 109-110*, 441-445.

(16) Johnson, H., Kenley, R. A. and Rynard, C. (1985) QSAR for cholinesterase inhibition by organophosphorous esters and CNDO/2 calculations for organophosphorous ester hydrolysis. *Quant. Struct.-Act. Relat. 4*, 172–180.

(17) Vighi, M., Garlanda, M. M. and Calamari, D. (1991) QSARs for toxicity of organophosphorous pesticides to *Daphnia* and honeybees. *Sci. Total Environ. 109-110*, 605–622.

(18) Worgan, A. D. P., Dearden, J. C., Edwards, R., Netzeva, T. I. and Cronin, M. T. D. (2003) Evaluation of a novel short-term algal toxicity assay by the development of QSARs and inter-species relationships for narcotic chemicals. *QSAR & Combinat. Sci. 22*, 204-209.

(19) Dimitrov, S., Koleva, Y., Schultz, T. W., Walker, J. D. and Mekenyan, O. (2004) Interspecies quantitative structure-activity relationship model for aldehydes: aquatic toxicity. *Environ Toxicol Chem 23*, 463-470.

(20) Cronin, M. T. D., Dearden, J. C. and Dobbs, A. J. (1991) QSAR studies of comparative toxicity in aquatic organisms. *The Science of The Total Environment 109-110*, 431-439.

(21) Cnubben, N. H., Peelen, S., Borst, J. W., Vervoort, J., Veeger, C. and Rietjens, I. M. (1994) Molecular orbital-based quantitative structure-activity relationship for the cytochrome P450-catalyzed 4-hydroxylation of halogenated anilines. *Chem. Res. Toxicol. 7*, 590-598.

(22) Rietjens, I. M. C. M., Soffers, A. E. M. F., Veeger, C. and Vervoort, J. (1993) Regioselectivity of cytochrome P-450 catalyzed hydroxylation of fluorobenzenes predicted by calculated frontier orbital substrate characteristics. *Biochemistry 32*, 4801-4812.

(23) OECD. (2004) OECD guidelines for testing of chemicals, guideline no. 202: *Daphnia* sp,. Acute Immobilization Test, pp 1-12, Paris, France.

(24) Elendt, B. P. (1990) Selenium deficiency in Crustacea - An ultrastructural approach to antennal damage in *Daphnia magna straus*. *Protoplasma 154*, 25-33.

(25) Mayer Jr, F. L. and Ellersieck, M. R. (1986) Manual of acute toxicity: interpretation and data base for 410 chemicals and 66 species of freshwater animals. *Resource Publication - US Fish & Wildlife Service 160*, 506.

(26) Bermudez-Saldana, J. M., Escuder-Gilabert, L., Medina-Hernandez, M. J., Villanueva-Camanas, R. M. and Sagrado, S. (2005) Chromatographic evaluation of the toxicity in fish of pesticides. *J. Chromatogr. B 814*, 115-125.

(27) Johnson, W. W. and Finley, M. T. (1980) Handbook of acute toxicity of chemicals to fish and aquatic invertebrates. *U.S. Fish Widl. Serv. Resour. Pub. 137*, 1–98.

(28) Tropsha, A., Gramatica, P. and Gombar, V. K. (2003) The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR and Combinatorial Science 22*, 69–77.

(29) Cronin, M. T. D. and Schultz, T. W. (2003) Pitfalls in QSAR. *Journal of Molecular Structure: THEOCHEM 622*, 39-51.

(30) Leo, A. (1993) Calculating log Poct from structures. *Chem. Rev. 93*, 1281-1306.

(31) Zvinavashe, E., Murk, A. J., Vervoort, J., Soffers, A. E. M. F., Freidig, A. and Rietjens, I. M. C. M. (2006) Quantum chemistry based quantitative structure-activity relationships for modeling the (sub)acute toxicity of substituted mononitrobenzenes in aquatic systems. *Environ. Toxicol. Chem. 25*, 2313–2321.

(32) Weininger, D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules *J. Chem. Inf. Comput. Sci. 28*, 31-36.

(33) Gramatica, P. (2007) Principles of QSAR models validation: Internal and external. *QSAR and Combinatorial Science 26*, 694-701.

(34) Zvinavashe, E., Berg, H. v. d., Soffers, A. E. M. F., Vervoort, J., Freidig, A., Murk, A. J. and Rietjens, I. M. C. M. (2008) QSAR models for predicting *in vivo* aquatic toxicity of chlorinated alkanes to fish. *Chem. Res. Toxicol. 21*, 739–745.

(35) Netzeva, T. I., Pavan, M. and Worth, A. P. (2008) Review of (quantitative) structure - Activity relationships for acute aquatic toxicity. *QSAR and Combinatorial Science 27*, 77-90.

(36) Petrauskas, A. A. and Kolovanov, E. A. (2000) ACD/Log P method description. *Perspect. Drug Discovery Des. 19*, 99–116.

(37) Worth, A. P. and Cronin, M. T. D. (2004) Report of the workshop on the validation of QSARs and other computational prediction models. *ATLA Alternatives to Laboratory Animals 32*, 703–706.

(38) Hutchinson, T. H., Barrett, S., Buzby, M., Constable, D., Hartmann, A., Hayes, E., Huggett, D., Laenge, R., Lillicrap, A. D., Straub, J. O. and Thompson, R. S. (2003) A strategy to reduce the numbers of fish used in acute ecotoxicity testing of pharmaceuticals. *Environ. Toxicol. Chem. 22*, 3031-3036.

(39) Hoekzema, C. C., Murk, A. J., Van De Waart, B. J., Van Der Hoeven, J. C. M. and De Roode, D. F. (2006) Alternative approaches can greatly reduce the number of fish used for acute toxicity testing. *Environ. Toxicol. Chem. 25*, 1322-1325.

On the number of EINECS compounds that can be covered by (Q)SAR models for acute toxicity

Elton Zvinavashe, Albertinka J. Murk, and Ivonne M.C.M. Rietjens

Toxicology letters ***(in press)***

Abstract

The new EU legislation for managing chemicals called REACH aims to fill in gaps in toxicity information that exist for the chemicals listed on the European Inventory of Existing Chemical Substances (EINECS). REACH advocates the use of alternatives to animal experimentation including, amongst others, (quantitative) structure-activity relationship models [(Q)SARs] to help fill in the toxicity data gaps. The aim of the present chapter was to provide a science-based estimate of the number of EINECS compounds that can be covered by (Q)SAR models for acute toxicity. Using the ECOSAR software, 54 % of the 100 196 EINECS chemicals were classified into 49 classes that can be potentially covered by (Q)SAR models. The largest proportion of the classified compounds (40 % of the EINECS list) falls into the classes of non-polar and polar narcotics. Compounds that were not classified include, for example, fish oils, botanical and animal extracts, and crude oil distillates. With rapid improvements in analytical tools, the number of EINECS compounds for which toxicity evaluations may be based on (Q)SAR approaches may be extended by further developing the method recently developed for the safety assessment of natural flavor complexes used as ingredients in food. This method is based on identification of the individual components in a mixture, and judgment of the safety of these identified individual compounds using toxicity information on structurally similar congeners in the respective classes. Such (Q)SAR approaches may be applied to an additional 2 938 EINECS compounds, representing botanical and animal extracts, leading to a total estimate of 57 % of the EINECS compounds for which (Q) SAR based approaches may assist in their safety assessment. It is concluded that, despite the fact that individual (Q)SARs may often each cover only a limited number, i.e. less than 1 %, of the EINECS compounds, the potential for applying (Q)SAR approaches for safety assessment of EINECS compounds may prove to be significant.

Introduction

The European Inventory of Existing Chemical Substances (EINECS) (*1*) contains more than 100 000 chemicals that were on the European Union (EU) market before 18 September 1981. After this date, about 4 300 new chemicals were introduced on the EU market and these compounds are listed on the European List of Notified Chemical Substances (*2*). It is recognized that insufficient (eco)toxicological information exists on the hazardous properties of many of the EINECS chemicals. This is mainly because for EINECS chemicals, unlike for ELINCS chemicals, there were no requirements for premarket testing, even though EINECS chemicals constituted 99 % of the total volume of all substances on the market at that time (*3*). To close the existing toxicity data gaps and to ensure better protection of human health and the environment from the risks that can be posed by chemicals, the EU adopted a new legislation called Registration, Evaluation, and Authorization of Chemicals (REACH) (*4*). This legislation came into force on June 1 2007 and shifts the responsibility of assessing and managing the risks posed by chemicals from governments to industry. In the REACH legislation, the amount of toxicity data required for a chemical increases with an increase in its production volume. As a result, concerns were expressed that for high production volume chemicals, large numbers of experimental animals might be required for toxicity testing in order to generate the required toxicity data. Therefore, within REACH, there is a provision to use, amongst others, sufficiently validated computational prediction models based on (quantitative) structure-activity relationships [(Q)SARs] to fill in the toxicity data gaps. This is expected to reduce the number of experimental animals used, and also to save time and costs (*4*).

Internationally, considerable efforts are in progress to help increase the acceptability of (Q)SAR models for regulatory purposes. In 2004, the Organization for Economic Cooperation and Development (OECD) proposed guidelines for QSAR model development and validation (*5*), that were recently updated (*6*). The OECD is also developing a software package called (Q)SAR Application Toolbox that member countries can use to fill toxicity gaps in (eco)toxicity data needed for assessing the hazards of chemicals (*7*). The European Chemicals Bureau (ECB) is currently compiling an inventory of QSAR models that can be used for regulatory purposes. For this purpose, they are developing a harmonized template, (Q)SAR Prediction Reporting Format, which will be used to summarize and report substance-specific predictions generated by (Q)SAR models (*8*).

An insight into the total number of EINECS compounds that QSAR models can potentially cover would be of interest for a potential successful application of these models

within REACH. In our previous studies, we developed validated QSAR models for three chemical groups, including mononitrobenzenes (chapter 2), chlorinated alkanes (chapter 3), and organothiophosphates (chapter 4). In these chapters we also identified how many of the EINECS chemicals would actually match the applicability domain of the respective QSARs. This analysis revealed that the QSARs for the mononitrobenzenes, chlorinated alkanes, and organothiophosphates covered 0.5 %, 0.1 %, and 0.1 % of the EINECS compounds respectively, together amounting to 660 compounds, making up about 0.7 % of the EINECS list. These numbers are schematically presented in Figure 5.1 and the question that emerges is to what extent (Q)SAR approaches can ultimately be expected to cover the more than 100 000 EINECS chemicals.

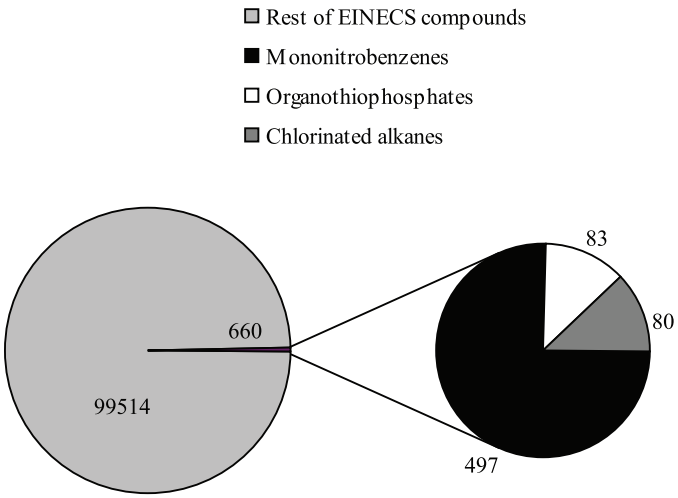


Figure 5.1: Proportion of EINECS chemicals covered by QSAR models we developed for three chemical groups including mononitrobenzenes (*9*), organothiophosphates (*10*), and chlorinated alkanes (*11*).

Therefore, the aim of the present study was to provide a science-based estimate of the number of EINECS compounds that can be covered by (Q)SAR models for toxicity. This was investigated by using the Ecological Structure Activity Relationships (ECOSAR) software, developed by the United States Environmental Protection Agency (US EPA), and quantifying to what extent EINECS chemicals can be classified into classes that can potentially be covered by these ECOSAR QSAR models. The effort also provided insight into what percentage and what type of EINECS chemicals would not qualify to be included in (Q)SAR type approaches and into the reasons why that would be the case.

MATERIALS AND METHODS

EINECS list

The EINECS list, containing the compounds to be classified was supplied by the ECB as a Microsoft® Excel file. It contains 100 196 compounds and for each compound, the following information is included: name, chemical abstracts service (CAS) number, EINECS number, and molecular formula. The CAS numbers were copied from the Excel file and saved into a Microsoft® Notepad (Microsoft, Redmond, WA, USA) file. The Notepad file was used as the input file into the ECOSAR class software.

Software

ECOSAR is a user-friendly computer programme developed and routinely applied by the US EPA for predicting aquatic toxicity to fish, daphnids and algae (*12*). The software version (0.99g) used in this study can be freely downloaded from the EPA website (*13*). From the main window of the ECOSAR software, the Microsoft® Notepad file was imported and processed in batch mode. The batch output file was set to tab delimited mode to allow easy export to Microsoft® Excel. The batch output file contained the name of the compound, CAS number, and QSAR class and it was imported into Microsoft® Excel 2003. The CAS numbers and compound names were compared with those in the EINECS list file for correctness.
Within Microsoft® Excel 2003 the numbers of compounds within each ECOSAR class were counted using a 15-day trial version of the software DigDB version 7.1.3.3 (Data Instrument Group, CA, USA; http://www.digdb.com). During installation, DigDB integrates into Microsoft® Excel as a plug-in. Using a "roll-up" option within DigDB, the compounds in each class were counted, and the results displayed in a new worksheet. The classes identified by ECOSAR were further grouped into mechanism-based classes according to the classification scheme proposed by Verhaar et al (*14*)

RESULTS

ECOSAR classification of EINECS compounds

Table 5.1 presents the results of classifying the 100 196 EINECS list compounds into the classes of the ECOSAR software. The ECOSAR software classified 54 277 (54%) out of the 100 196 EINECS compounds into 49 classes. For compounds containing more than one functional group, ECOSAR lists all the functional groups that it can identify in a molecule and

may assign more than one functional class to a specific molecule. This is illustrated in Figure 5.2 for the compound 2,5-dihydroxy-3,6-dichlorobenzoquinone; (CAS 87-88-7), where ECOSAR identifies three functional groups and classifies the compound into the class vinyl/allyl halides and identifies two additional classes namely quinone/hydroquinone and vinyl/allyl alcohols.

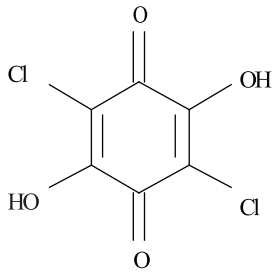


Figure 5.2: 2,5-dihydroxy-3,6-dichlorobenzoquinone (CAS 87-88-7), an example of a compound that is classified into the classes vinyl/allyl halides, quinone/hydroquinone and vinyl/allyl alcohols by ECOSAR due to the presence of several functional groups.

Table 5.1, therefore, also includes information on which additional functional groups were found to occur concomitantly at least in some congeners in the respective class and lists in column 3 the respective classes corresponding to these additional functional groups encountered in a specific class. The results thus obtained reveal that neutral organics form the largest class (class 1, n = 21 233). The ECOSAR software defines neutral organics as compounds that are generally solvents, non-ionizable, and non-reactive. Table 5.1 also shows that a large majority (80 %) of the classified compounds fall into seven classes: neutral organics, esters, aliphatic amines, phenols and aromatic amines, acrylates and aldehydes (classes 1 to 7). The remaining compounds (20 %) fall into 42 classes that contain less than 1,000 compounds per class (Table 5.1; classes 8 to 49).

Further grouping of ECOSAR classes

Verhaar et al. developed a classification scheme for environmental pollutants *(14)* in which chemicals are divided into four categories: i) non-polar narcotics, ii) polar narcotics, iii) chemicals with unspecific reactivity, and (iv) chemicals with a specific mechanism of action. Non-polar narcotics include aliphatic and aromatic hydrocarbons, halogenated aliphatic and aromatic hydrocarbons, ethers, alcohols and others (*14*). Considering the ECOSAR classification scheme (Table 5.1), the neutral organics class (ECOSAR class 1), which contains the largest number of classified compounds, would fall into the non-polar narcotics group.

Table 5.1: Numbers of compounds in various classes as identified by the ECOSAR software.

Class number	Class name [a]	Other functional groups attached to the main class identified by the ECOSAR software	Total
1	neutral organics	carbon, hydrogen, bromine, phosphorous, iodine, chlorine, fluorine, selenium, sulphur, oxygen, nitrogen, boron, silicon, germanium, arsenic, antimony, tellurium, polonium, alkali, alkaline, lanthanides, actinides, transition metals	21233
2	esters	phosphate esters, vinyl/allyl ketones, surfactants-anionic, phenols, salicylates	8259
3	aliphatic amines	esters, phenols, benzyl alcohols, benzyl amines	5985
4	phenols	salicylic acid, quinone/hydroquinone, benzyl alcohols, triazines	3561
5	aromatic amines	phenols, esters, imidazoles, triazines	2581
6	acrylates	esters, aliphatic amines, vinyl/allyl alcohols, allylic/vinyl nitriles	1253
7	aldehydes	phenols , esters	1032
8	hydrazines	phenols, imides, triazines	920
9	imides	imidazoles, phenols, vinyl/allyl halides , vinyl/allyl alcohols	852
10	vinyl/allyl ketones	vinyl/allyl alcohols, vinyl/allyl ethers	831
11	acid halide	benzyl halides, thiophenes, vinyl allyls	606
12	surfactants-anionic	vinyl/allyl alcohols, imidazoles, vinyl/allyl ethers	515
13	surfactants-cationic	surfactants-anionic, triazines	493
14	benzyl alcohols	imidazoles, benzyl halides	462
15	vinyl/allyl ethers	vinyl/allyl alcohols , thiazolidinones, quinone/ hydroquinone	430
16	epoxides	esters, vinyl/allyl ketones, silanes, vinyl/allyl ethers	392
17	benzyl halides	imidazoles	369
18	imidazoles	no other substituents	349
19	methacrylates	silanes	341
20	dinitrobenzenes	esters, phenols, hydrazines	340
21	vinyl/allyl alcohols	no other substituents	283
22	silanes alkoxy	benzyl halides	260

[a] *class name as assigned by the ECOSAR software.*

Class number	Class name [a]	Other functional groups attached to the main class identified by the ECOSAR software	Total
23	anilines (amino-meta)	amino-para anilines, amino-ortho anilines, triazines	252
24	vinyl/allyl halides	quinone/hydroquinone , vinyl/allyl alcohols, acrylamides	240
25	peroxy acids	thiophenes, triazines, vinyl/allyl ethers	199
26	triazines	vinyl/allyl ethers	198
27	thiols–mercaptans	silanes	193
28	isocyanates	esters, imides	183
29	acrylamide	vinyl/allyl alcohols	183
30	allylic/vinyl nitriles	acrylamides, vinyl/allyl alcohols , vinyl/allyl ethers	174
31	diazoniums	esters	167
32	thiophenes	benzyl halides , acrylamides, vinyl/allyl alcohols	150
33	schiff bases	vinyl/allyl ketones , vinyl/allyl ethers	127
34	haloacetamides	benzyl halides	125
35	diepoxides	esters	118
36	diketones	esters, vinyl/allyl ketones	95
37	thiocyanates	silamines	69
38	surfactants-nonionic	silanes	65
39	thiazolidinones	vinyl/allyl halides	62
40	silamines	no other substituents	58
41	benzotriazoles	phenols	57
42	aziridines	esters	38
43	quinone/ hydroquinone	vinyl/allyl alcohols	37
44	benzyl amines	imidazoles	29
45	propargyl alcohols	propargyl ethers, propargyl amines, vinyl/allyl halides	26
46	propargyl ethers	no other substituents	23
47	dinitro aromatic amine	phenols, thiophenes , diazoniums	23
48	vinyl/allyl sulfones	vinyl/allyl ethers , acrylamides	21
49	propargyl amines	no other substituents	18
	Total		54277

[a] class name as assigned by the ECOSAR software.

Aromatic amines (ECOSAR class 2), esters (ECOSAR class 25), phenols (ECOSAR class 26), and anilines (ECOSAR class 35) would fit into the polar narcotics category (*14*). Based on these considerations it can be concluded that, as schematically represented in Figure 5.3, the polar narcotics would include 17 077 EINECS list chemicals (ECOSAR class 2+25+26+35) and the non-polar narcotics would include 21 233 EINECS compounds (ECOSAR class 1), together covering about 40 % of the total EINECS list.

Compounds in Table 5.1 that fit the unspecific reactivity category of Verhaar et al. include propargyls (classes 16–18), cyanates (classes 19 and 20), aldehydes (class 28), hydrazines (class 29), and acid halides (class 31). The ECOSAR software classifies reactive compounds (e.g. organophosphorous, pyrethroids, carbamates etc.) into the ester class (ECOSAR class 25) since they are esterified.

Compounds not classified by ECOSAR

The ECOSAR software was not able to classify the remaining 45 919 EINECS compounds. Further analysis of these compounds revealed that 18 367 (40 %) of them do not have a defined molecular formula. These include for example fish oils (e.g. cod liver oil; CAS 8001-69-2), plant oils (e.g. rape oil monoglycerides; CAS 85586-30-7), botanical extracts (e.g. sweet pea extract; CAS 90604-48-1), animal extracts (civet secretion; CAS 68916-26-7), crude oil distillates (e.g. low temperature crude tar bases; CAS 141785-66-2), enzymes (e.g. luciferase; CAS 9014-00-0), and wastes (e.g. ethylene oxide absorber reactor waste gases; CAS 68513-74-6). Botanical extracts (n = 2 819) and animal extracts (n = 119) in total (n = 2 938) account for 16 % of the compounds without a defined molecular formula.

The remaining 27 552 compounds which ECOSAR was unable to classify fall into the groups of for example, medicines (e.g. streptomycin B; CAS 128-45-0, morphine hydrochloride, CAS 52-26-6) and salts (e.g. formic acid lead salt; CAS 7056-83-9).

Discussion

The REACH legislation focuses on providing toxicity information for about 30 000 EINECS chemicals with an annual production volume greater than 1 ton (*4*). The rationale for focusing on production volume is that large production volumes increase the potential for exposure, and therefore the potential risks associated with the chemical (*15*). In the present study, we chose to focus on the complete EINECS list, rather than being restricted by production volumes. This is because it is not known to what extent production volume actually predicts exposure, and chemicals with low toxicity may be overrepresented among high volume chemicals. Therefore, there can be a bias towards less acutely toxic chemicals when one would only consider the 30 000 chemicals with annual production volumes greater than 1 ton (*15*). Equally important is the fact that during development of QSAR models, toxicity data from both low and high production volume chemicals is often used, so one might also expect to predict toxicity of both low and high volume chemicals by these QSARs.

In order to implement QSAR models successfully, one needs to be able to classify compounds into specific groups, based on chemical structure or mechanism of action (*16*). In the present study, using the ECOSAR software, we classified approximately 54 % of the 100 196 EINECS chemicals into 49 classes (Table 5.1) that can be potentially covered by QSAR models. As far as we are aware, this is the first attempt to categorize the complete EINECS list into specific classes. The usefulness of ECOSAR for classification of chemicals has been previously demonstrated. Hulzebos and Posthumus used ECOSAR to categorize 70 chemicals into chemical classes with 87 % accuracy, and this compared well (90 % accuracy) with using DEREK, a commercial software package (*17*). Thus, the ECOSAR software, which is freely available for download on the internet, seems an appropriate method for chemical classification.

Using the classification scheme proposed previously by Verhaar et al. (1992), we further grouped the ECOSAR classes into four groups. It has been frequently demonstrated that the acute toxicity of both non-polar and polar narcotics can be adequately explained by their hydrophobicity, as they are assumed to exert their toxicity via a non-specific mode of action (*14, 18-20*). Most of the QSAR models published over the last 30 years have focused on both non-polar and polar narcotics and describe QSARs for example, for chlorobenzenes (*21*), alcohols and chlorohydrocarbons (*22*), halogenated aliphatics (*23*), chlorinated alkanes (*11, 18, 24*). Considering that the non-polar and polar narcotics (Fig. 5.3; total of $n = 38\,310$) cover about 40 % of the EINECS compounds, the potential application of QSAR models for these compounds looks promising.

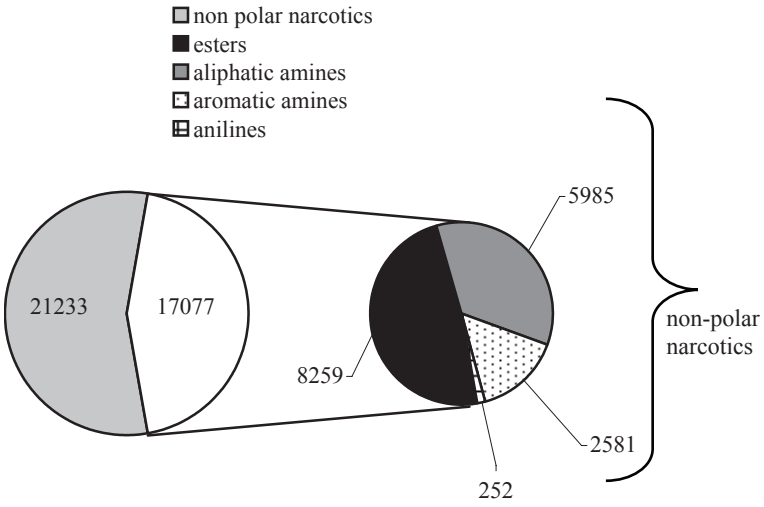


Figure 5.3: Further grouping of ECOSAR classes into non-polar and polar narcotics according to the classification scheme proposed by Verhaar et al (*14*). The polar narcotics group (smaller pie chart) is made up of esters, aliphatic amines, aromatic amines, and anilines.

However, one needs to be aware that even within each of the specific classes, a single QSAR model would probably not cover all of the compounds. This is illustrated in Figure 5.4, showing that a QSAR model developed for substituted mononitrobenzenes could not cover substituted nitrobenzenes that (i) are ionized at physiological pH of 7.4, i.e. nitrophenols and nitrobenzoic acids (□), (ii) have two nitro groups (Δ), (iii) contain a cyano substituent (O), or (iv) have two phenyl rings (◊), because they were clearly outliers in the model probably due to a different mechanism of action (*9*).

The ECOSAR software classifies nitrophenols (ECOSAR class 26) and dinitrobenzenes (ECOSAR class 33) into different classes than the substituted mononitrobenzenes (ECOSAR class 1), thus confirming their outlier status. Although nitrobenzoic acids and cyanonitrobenzenes fall into the same ECOSAR class as the substituted mononitrobenzenes, (ECOSAR class 1), other QSAR models would need to be defined for them.

The ECOSAR software does not separately categorize specifically reactive compounds (e.g. organophosphorous, pyrethroids, carbamates etc). For example, organophosphorous compounds exert their specific acute toxic effects primarily by inhibiting the enzyme acetylcholinesterase (AChE) (*25, 26*) and pyrethroids interfere with the closing of sodium channels in neurons (*27*). Most of the specifically reactive compounds are used as pesticides and they have been the focus of several QSAR modeling efforts as demonstrated by

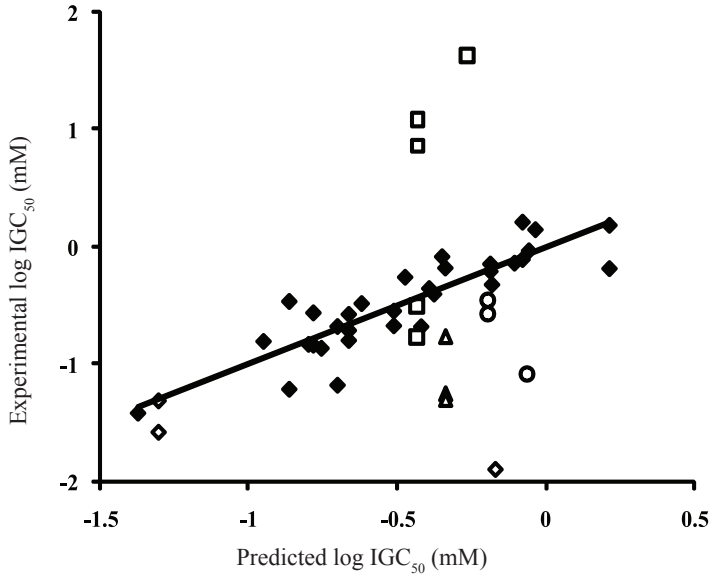


Figure 5.4: QSAR for the correlation between the decadic logarithm of the 48 h experimental toxicity of substituted mononitrobenzenes to the protozoan *Tetrahymena pyriformis* and the toxicity predicted on the basis of log K_{ow} for (r^2 = 0.815; n = 33), N.B. Data points not included represent substituted nitrobenzenes that (i) are ionized at physiological pH of 7.4, i.e. nitrophenols and nitrobenzoic acids (□), (ii) have two nitro groups (Δ), (iii) contain a cyano substituent (O), or (iv) have two phenyl rings (◊). $IGC_{50} = IC_{50}$ for growth inhibition. Adapted from literature (*9*).

QSARs described for e.g. organophosphorous compounds (*28-34*), pyrethroids (*35-38*), and carbamates (*31, 33, 34, 39*). Thus, it can be expected that compounds of ECOSAR classes that would fall into these groups of chemicals with a specific mechanism of action can all be covered by QSARs.

For some of the compounds that the ECOSAR software was not able to classify, it is clear that QSAR models would never apply to them. Substances that fall into the categories petroleum and coal based substances and waste gases cannot be tested with conventional test systems due to their complex and variable composition (*40*), making the application of QSAR models unrealistic. On the other hand, the safety assessment of botanical extracts could be conducted based on further development of an approach used for natural flavor ingredients in food (*41*). The approach developed for essential oils is based on the concept that the types of chemicals in natural flavor complexes are not infinite in structural variation, because chemical constituents in plants originate from a limited number of biosynthetic pathways (*41*). Therefore, if more than 95% of the essential oil constituents can be adequately characterized and assigned to well defined congeneric groups, the safety of the structural classes of

identified individual compounds can be judged using toxicity information on structurally similar congeners in the respective classes (*41*) in a SAR approach. Considering the rapid improvement and reduced costs of analytical tools, compared to the ever increasing costs for traditional toxicity testing, this approach may prove to be potentially useful and applicable for a wider range of botanical extracts (n = 2 819), and/or animal extracts (n = 119). This would lead to an increase in the coverage of the EINECS list by (Q)SAR approaches from 54 277 to 57 215 compounds, equivalent to 57 % of the EINECS list.

As classification of compounds is an important first step for the implementation of QSAR models on the regulatory scene, a next step would be to evaluate the quality of the existing QSAR models for the identified classes, a process that is currently ongoing at the ECB (*8*). Altogether, the results of the present study reveal that, despite the fact that individual QSARs may often each cover only limited, i.e. less than 1%, of the EINECS compounds, the potential for applying (Q)SAR approaches for safety assessment of EINECS compounds may prove to be significant.

ACKNOWLEDGEMENTS

We are grateful to the Netherlands Organization for Health Research and Development, (ZonMw; project no. 3170.0066) for funding this study under the theme “Alternatives to animal experimentation”.

References

(1) EU. (1993) Council Regulation (EEC) No 793/93 of 23 March 1993 on the evaluation and control of the risks of existing substances *Off. J. Eur. Union. L 084*, 1–75.

(2) ELINCS. European List of Notified Chemical Substances; available online at http://ecb.jrc.it/esis/index.php?PGM=eli). .

(3) EU. (2001) White Paper: Strategy for a future chemicals policy. Commission of the European Committees, Brussels, Belgium. COM(2001) 88 final., pp 1–32.

(4) EU. (2006) Regulation (EC) No 1907/2006 of the European Parliament and of the Council. *Off. J. Eur. Union*, 1–849.

(5) OECD. (2004) The report from the expert group on (Quantitative) Structure-Activity Relationship ((Q)SARs) on the principles for the validation of (Q)SARS. ENV/JM/MONO(2004)24. , In *OECD Series on testing and assessment: number 49*.

(6) OECD. (2007) Guidance document on the validation of (Quantitative) structure-activity relationships [(Q)SAR] models, ENV/JM/MONO(2007)2, In *OECD Environment Health and Safety Publications: Series on Testing and Assessment No. 69. Paris, France* pp 1–154.

(7) OECD. (Q)SAR application toolbox; available online at http://www.oecd.org/document/23/0,3343,en_2649_34379_33957015_1_1_1_1,00.html

(8) ECB. (Q)SAR Prediction Reporting Format; available online at http://ecb.jrc.it/qsar/qsar-tools/index.php?c=QRF.

(9) Zvinavashe, E., Murk, A. J., Vervoort, J., Soffers, A. E. M. F., Freidig, A. and Rietjens, I. M. C. M. (2006) Quantum chemistry based quantitative structure-activity relationships for modeling the (sub)acute toxicity of substituted mononitrobenzenes in aquatic systems. *Environ. Toxicol. Chem. 25*, 2313–2321.

(10) Zvinavashe, E., Tingting, D., Griff, T., Berg van den, H., Soffers, A. E. M. F., Vervoort, J., Freidig, A., Murk, A. J. and Rietjens, I. M. C. M. (2008) Quantitative structure-activity relationship modeling of the toxicity of organothiophosphate pesticides to aquatic organisms and honeybee. *Submitted*.

(11) Zvinavashe, E., Berg, H. v. d., Soffers, A. E. M. F., Vervoort, J., Freidig, A., Murk, A. J. and Rietjens, I. M. C. M. (2008) QSAR models for predicting *in vivo* aquatic toxicity of chlorinated alkanes to fish. *Chem. Res. Toxicol. 21*, 739–745.

(12) Nabholz, J. V., Clements, R. G., Zeeman, M. G., Osborn, K. C. and Wedge, R. (1993) Validation of structure activity relationships used by the USEPA's Office of Pollution Prevention and Toxics for the environmental hazard assessment of industrial chemicals, In *ASTM Special Technical Publication* pp 571–590.

(13) EPA. ECOSAR software version 0.99g; available online at http://www.epa.gov/oppt/newchems/tools/21ecosar.htm.

(14) Verhaar, H. J. M., van Leeuwen, C. J. and Hermens, J. L. M. (1992) Classifying environmental pollutants. *Chemosphere 25*, 471-491.

(15) Hansson, S. O. and Ruden, C. (2006) Priority setting in the REACH system. *Toxicol. Sci. 90*, 304–348.

(16) Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., McDowell, R. M. and Gramatica, P. (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ. Health Perspect. 111*, 1361–1375.

(17) Hulzebos, E. M. and Posthumus, R. (2003) (Q)SARs: Gatekeepers against risk on chemicals? *SAR and QSAR Environ. Res. 14*, 285–316.

(18) Könemann, H. (1981) Quantitative structure-activity relationships in fish toxicity studies Part 1: Relationship for 50 industrial pollutants. *Toxicology 19*, 209–221.

(19) Veith, G. D. and Broderius, S. J. (1990) Rules for distinguishing toxicants that cause type I

and type II narcosis syndromes. *Environ. Health Persp. 87*, 207–211.

(20) Veith, G. D., Call, D. J. and Brooke, L. T. (1983) Structure - toxicity relationships for the fathead minnow, *Pimephales promelas*: narcotic industrial chemicals. *Can. J. Fish. Aquat. Sci. 40*, 743–748.

(21) Calamari, D., Galassi S., Setti F. and Vighi M. (1983) Toxicity of selected chlorobenzenes to aquatic organisms. *Chemosphere 12*, 253–262.

(22) Hermens, J., Broekhuyzen, E., Canton, H. and Wegman, R. (1985) Quantitative structure-activity relationships and mixture toxicity studies of alcohols and chlorohydrocarbons: Effects on growth of *Daphnia magna*. *Aquatic Toxicology 6*, 209–217.

(23) Akers, K. S., Sinks, G. D. and Schultz, T. W. (1999) Structure-toxicity relationships for selected halogenated aliphatic chemicals. *Environ. Toxicol. Pharmacol. 7*, 33–39.

(24) Russom, C. L., Bradbury, S. P., Broderius, S. J., Hammermeister, D. E. and Drummond, R. A. (1997) Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). *Environ. Toxicol. Chem. 16*, 948–967.

(25) Vittozzi, L., Fabrizi, L., Di Consiglio, E. and Testai, E. (2001) Mechanistic aspects of organophosphorothionate toxicity in fish and humans. *Environ. Int. 26*, 125–129.

(26) Nigg, H. N. and Knaak, J. B. (2000) Blood cholinesterases as human biomarkers of organophosphorous pesticide exposure. *Rev. Environ. Contam. Toxicol. 163*, 29–111.

(27) Narahashi, T., Zhao, X., Ikeda, T., Nagata, K. and Yeh, J. Z. (2007) Differential actions of insecticides on target sites: Basis for selective toxicity. *Human Experim. Toxicol. 26*, 361–366.

(28) Vighi, M., Garlanda, M. M. and Calamari, D. (1991) QSARs for toxicity of organophosphorous pesticides to *Daphnia* and honeybees. *Sci. Total Environ. 109-110*, 605–622.

(29) Galli, R., Rich, H. W. and Scholtz, R. (1994) Toxicity of organophosphate insecticides and their metabolites to the water flea Daphnia magna, the Microtox test and an acetylcholinesterase inhibition test. *Aquat. Toxicol. 30*, 259–269.

(30) Fukuto, T. R. (1990) Mechanism of action of organophosphorous and carbamate insecticides. *Environ. Health Perspect. 87*, 245–254.

(31) Bermudez-Saldana, J. M. and Cronin, M. T. D. (2006) Quantitative structure-activity relationships for the toxicity of organophosphorous and carbamate pesticides to the rainbow trout *Onchorhyncus mykiss*. *Pest Manag. Sci. 62*, 819–831.

(32) Johnson, H., Kenley, R. A. and Rynard, C. (1985) QSAR for cholinesterase inhibition by organophosphorous esters and CNDO/2 calculations for organophosphorous ester hydrolysis. *Quant. Struct.-Act. Relat. 4*, 172–180.

(33) Devillers, J. (2001) A general QSAR model for predicting the acute toxicity of pesticides to *Lepomis macrochirus*. *SAR and QSAR Environ. Res. 11*, 397–417.

(34) Devillers, J. and Flatin, J. (2000) A general QSAR model for predicting the acute toxicity of pesticides to *Oncorhynchus mykiss*. *SAR and QSAR Environ. Res. 11*, 25–43.

(35) Ford, M. G., Hoare, N. E., Hudson, B. D., Nevell, T. G. and Banting, L. (2002) QSAR studies of the pyrethroid insecticides: Part 3. A putative pharmacophore derived using methodology based on molecular dynamics and hierarchical cluster analysis. *Journal of Molecular Graphics and Modelling 21*, 29-36.

(36) Chuman, H., Goto, S., Karasawa, M., Sasaki, M., Nagashima, U., Nishimura, K. and Fujita, T. (2000) Three-dimensional structure-activity relationships of synthetic pyrethroids: 2. Three-dimensional and classical QSAR studies. *Quant. Struct.-Act. Relat. 19*, 455–466.

(37) Mathew, N., Subramanian, S. and Kalyanasundaram, M. (1996) Quantitative structure activity relationship studies in pyrethroid esters derived from substituted 3-methyl-2-phenoxybutanoic acids against Culex quinquefasciatus. *Indian Journal of Chemistry - Section B Organic and Medicinal Chemistry 35*, 40-44.

(38) Hudson, B. D., George, A. R., Ford, M. G. and Livingstone, D. J. (1992) Structure-activity relationships of pyrethroid insecticides. Part 2. The use of molecular dynamics for conformation searching and average parameter calculation. *Journal of Computer-Aided Molecular Design 6*, 191-201.

(39) Lin, G., Liao, W. C., Chan, C. H., Wu, Y. H., Tsai, H. J. and Hsieh, C. W. (2004) Quantitative structure-activity relationships for the pre-steady state acetylcholinesterase inhibition by carbamates. *J. Biochem. Mol. Toxicol. 18*, 353–360.

(40) Allanou, R., Hansen, B. G. and van der Bilt, Y. (1999) Public availability of data on EU high production volume chemicals, pp 1–21, European Commission, JRC, Institute for Health and Consumer Protection, European Chemicals Bureau, Italy.

(41) Smith, R. L., Cohen, S. M., Doull, J., Feron, V. J., Goodman, J. I., Marnett, L. J., Portoghese, P. S., Waddell, W. J., Wagner, B. M., Hall, R. L., Higley, N. A., Lucas-Gavin, C. and Adams, T. B. (2005) A procedure for the safety evaluation of natural flavor complexes used as ingredients in food: Essential oils. *Food Chem. Toxicol. 43*, 345–363.

Summary, general discussion and future perspectives

partly based on:

The promises and pitfalls of QSAR approaches for predicting metabolism and toxicity

Elton Zvinavashe, Albertinka J. Murk and Ivonne M.C.M. Rietjens

Chemical Research in Toxicology ***(in press)***

The new EU legislation for managing chemicals called Registration, Evaluation and Authorization of Chemicals (REACH) (*1*) aims to fill in gaps in toxicity information that exist for the chemicals listed on the European Inventory of Existing Chemical Substances (EINECS) (*2*). Within the REACH framework, manufacturers are required by the year 2018 to provide toxicity information for about 30 000 chemicals with annual production volumes greater than 1 ton (*1*). The amount of toxicity data required for a chemical increases with an increase in its production volume. There is currently ongoing debate about the potentially large number of animals that have to be used for experimental toxicity testing as an outcome of REACH. In 2005, about 12 million animals were used within the EU (*3*). Within the REACH framework, there are provisions to allow alternative testing methods e.g. *in vitro* and *in silico* approaches. Due to the large amount of toxicity information that is required within the next 10 years, methods that can help to direct priorities for future toxicity testing will help to save time, money, and animals. In this regard, *in silico* methods such as those based on (Quantitative) Structure-Activity Relationships [(Q)SARs] are expected to provide toxicity information or identify hazards of chemicals, and this information can then be used to identify priority chemicals for further risk evaluation or even predict toxicity of congeners within the applicability domain of the QSAR.

A QSAR is based on the assumption that the biological activity of a new or untested chemical can be inferred from the molecular structure, or properties of similar compounds whose activities have already been assessed (*4*). Traditional QSARs use experimentally determined descriptors (*5, 6*). However, due to lack of large data sets of experimentally derived descriptors, QSARs approaches have been developed that are based on descriptors defined using quantum-mechanical computer calculations. Due to the increasing power of computers, quantum-mechanical computer calculations have become a valuable and widely applicable tool in (bio)chemical research. They provide the possibility to calculate, using the chemical structure of a compound as the sole input, parameters that define the physico-chemical properties and relative reactivity of a compound. Computational approaches offer the advantages of ease of use, speed, and low costs.

In 2007, the Organization for Economic Cooperation and Development (OECD) published a guidance document for (Q)SAR model development and validation in order to help increase the acceptability of (Q)SAR models for regulatory purposes (*7*). These guidelines state that in order to facilitate the consideration of a (Q)SAR model for regulatory purposes, the (Q)SAR model should be associated with the following five categories of information:1) a defined endpoint, 2) an unambiguous algorithm, 3) a defined domain of applicability, 4) appropriate measures of goodness-of-fit, robustness, and predictivity, and 5) a mechanistic

interpretation (*7*). The aim of this thesis was to develop computational chemistry-based QSAR approaches that enable identification of priorities within various selected groups of EINECS chemicals, and to investigate to what extent QSAR approaches can be of use in predicting toxicty of the large number of EINECS chemicals. To this end, validated QSAR models for acute toxicity of selected groups of EINECS chemicals were developed taking into account the OECD guidelines and the actual number of EINECS compounds covered by these QSAR approaches was established.

Summary of methods and results

In this thesis, QSAR models were developed for nitrobenzenes **(chapter 2)**, chlorinated alkanes **(chapter 3)**, and organothiophosphate pesticides **(chapter 4),** Three descriptors were investigated for their suitability in modeling the toxicity of the chemicals in the three groups to various organisms. These were;

(i) the logarithm of the octanol/water partition coefficient ($\log K_{ow}$), which generally models a compound's hydrophobicity, which is important in describing the passage of a compound through membranes,

(ii) the energy of the lowest unoccupied molecular orbital E_{lumo} (eV), which models the electrophilic nature of the chemicals, and

(iii) the energy of the highest occupied molecular orbital E_{homo} (eV), which models the nucleophilic nature of the chemicals.

In a final step, a science-based estimate was made of the percentage of EINECS chemicals that can be grouped into specific chemical classes, and thus in theory be subject to QSAR modeling **(chapter 5)**.

The initial task in this thesis was to identify suitable software packages, descriptor calculation protocols, and statistical techniques for use for the rest of the thesis. This was done using a group of compounds called substituted nitrobenzenes as they had sufficient experimental toxicity data from the literature for modeling purposes (**chapter 2**). Substituted nitrobenzenes are widely used in industry during the synthesis of dyes, explosives, solvents, plastics, anilines, and various bioactive products such as insecticides, pesticides, and pharmaceuticals (*8, 9*). As a result of their varied origins and uses, substituted nitrobenzenes are widespread in ecosystems and consequently have a high potential for causing ecotoxic effects (*9*). Out of six software packages, ClogP was identified as the most suitable one for estimating $\log K_{ow}$ values, a conclusion based on the highest correlation ($r^2 = 0.98$) between experimental and predicted $\log K_{ow}$ values for a set of substituted nitrobenzenes. Using experimental literature data sets on the acute toxicity of substituted nitrobenzenes to

algae, daphnids, fish, protozoans, bacteria, and yeast the possibilities to establish quantum chemistry-based QSARs were investigated. The log K_{ow} was a sufficient descriptor ($0.65 < r^2 < 0.98$) in explaining the toxicity of substituted mononitrobenzenes to protozoans, fish, daphnids, yeast, with an additional electronic descriptor, $E_{lumo,}$ being required for algae. The QSARs were valid for neutral substituted mononitrobenzenes with no -OH, -COOH, or -CN substituents attached directly to the ring. Based on these substituent criteria, 497 EINECS compounds were identified that fit the selection criteria for the established QSARs. The log K_{ow} based QSARs for the substituted mononitrobenzenes were compared to literature QSARs that describe the minimum or baseline toxicity of chemicals due to narcotic action in order to identify log K_{ow} cutoff points, above which one should apply the baseline QSAR models instead of the ones defined in our studies. Based on these comparisons, an advisory tool was developed that directs users to the appropriate QSAR model to apply for five types of organisms within specified log K_{ow} ranges.

In a next step, QSAR models were developed for chlorinated alkanes, a group of chemicals with large production volumes that are widely used as industrial and household solvents, fumigants, and intermediates in chemical synthesis (*10*), and therefore, have a large potential for environmental pollution. Whereas in c**hapter 2** there were sufficient experimental toxicity data for substituted mononitrobenzenes available in the literature for developing QSAR models, this was not the case for the chlorinated alkanes. Therefore, in **chapter 3**, Chinese hamster ovary (CHO) cells were exposed to varying concentrations of a series of chlorinated alkanes in an *in vitro* cytotoxicity assay in order to generate *in vitro* toxicity data that could be used to develop QSAR models and eventually be correlated to the limited available *in vivo* data to describe their acute *in vivo* toxicity. Cytotoxicity of the series of chlorinated alkanes to CHO cells was observed at concentrations similar to those that had been shown previously to be toxic to fish (*11*). Strong correlations were observed between the acute *in vitro* toxicity of the chlorinated alkanes and (i) log K_{ow} ($r^2 = 0.88$) and (ii) *in vivo* acute toxicity to fish ($r^2 = 0.76$). A QSAR model was developed to predict *in vivo* acute toxicity to fish based on the *in vitro* data and even on *in silico* log K_{ow} data only. The QSAR models were validated both internally and externally. The developed QSAR models are applicable to chlorinated alkanes with up to 10 C-atoms, up to eight Cl-atoms, and log K_{ow} values lying within the range from 1.71 to 5.70, and they cover 77 EINECS chemicals.

The QSAR methodologies applied to substituted mononitrobenzenes (**chapter 2**) and chlorinated alkanes (**chapter 3**) were extended to a third group of chemicals, organothiophosphates (**chapter 4**). Organothiophosphates are the most widely used pesticides since the banning of organochlorines in the 1970s (*12*) as they have lower environmental

persistence, but they are often implicated in wildlife and human poisonings (*12*). They exert their primary acute toxic effects by inhibiting the enzyme acetylcholinesterase (AChE). As a result AChE can no longer hydrolyze its natural substrate, the neurotransmitter acetylcholine, and acetylcholine accumulates at the synaptic terminals leading to overstimulation of the postsynaptic receptor (*13, 14*). In order to exert their full toxicity, organothiophosphates need to be oxidized to their active oxo-form form by cytochrome P_{450} enzymes (*15*). In this chapter, the descriptors investigated were log K_{ow}, E_{homo} (to model the oxidation of the organothiophosphate to its active oxo-form by cytochrome P_{450} enzymes), and E_{lumo} (to model the binding of the active oxo-form of the organothiophosphate to AChE). Toxicity data sets used were from (i) acute immobilization experiments conducted in the present thesis with *Daphnia magna* and from (ii) the literature describing toxicity to fish. Additionally, it was investigated if toxicity data for the invertebrate *D. magna* could be used to build a QSAR model to predict toxicity to fish. Suitable QSAR models ($0.80 < r^2 < 0.82$) were derived to predict acute toxicity of organothiophosphates to fish and *D. magna.* Toxicity data for *D. magna* correlated well ($r^2 = 0.94$) with toxicity data for fish. This implied that by performing toxicity tests with *D. magna*, one could use our interspecies QSAR model to predict the acute toxicity of organothiophosphates to fish. The three QSAR models were validated either both internally and externally (*D. magna*) or internally only (carp and *D. magna* to carp). For each QSAR model, an applicability domain was defined based on the chemical structures and the ranges of the descriptor values of the training set compounds. Eighty-three EINECS compounds were identified that fit the selection criteria for the QSAR models.

The QSAR models for the mononitrobenzenes (**chapter 2**), chlorinated alkanes (**chapter 3**), and organothiophosphates (**chapter 4**) covered respectively 0.5 %, 0.1 %, and 0.1 % of the 100 196 EINECS compounds respectively, together amounting to a total of 660 compounds, making up about 0.7 % of the EINECS list. As the three QSAR models appeared to cover only a small percentage of the EINECS list, in a next step, an estimate was made of the percentage of EINECS chemicals that can be grouped into specific chemical classes, and thus in theory be subject to QSAR modeling **(chapter 5).** To this end, ECOSAR, a software package that can be freely downloaded over the internet, was used to group the EINECS chemicals into various classes based on chemical structure. The ECOSAR software classified 54 % of the 100 196 EINECS chemicals into 49 classes that can be potentially covered by (Q)SAR models. The largest proportion of the classified compounds (40 % of the EINECS list) fall into the classes of non-polar and polar narcotics. Compounds that were not classified include, for example, fish oils, botanical and animal extracts, and crude oil distillates. We proposed that the safety assessment of botanical extracts could be done

by further development of a method recently reported for the safety assessment of natural flavour complexes used as ingredients in food (*16*). The approach developed for essential oils is based on the concept that the types of chemicals in natural flavour complexes are not infinite in structural variation, because chemical constituents in plants originate from a limited number of biosynthetic pathways (*16*). Therefore, if more than 95% of the essential oil constituents can be adequately characterized and assigned to well defined congeneric groups, the safety of the structural classes of identified individual compounds can be judged using toxicity information on structurally similar congeners in the respective classes (*16*) in a SAR approach. This would result in an additional 3 % of the EINECS compounds that could be covered by SAR approaches, bringing the total percentage of EINECS compounds that can be covered by (Q)SAR approaches to 57 %. A summary of the QSAR models for acute toxicity developed in this thesis, their applicability domains and the organisms for which the QSAR models can acute toxicity predictions are shown in Figure 6.1

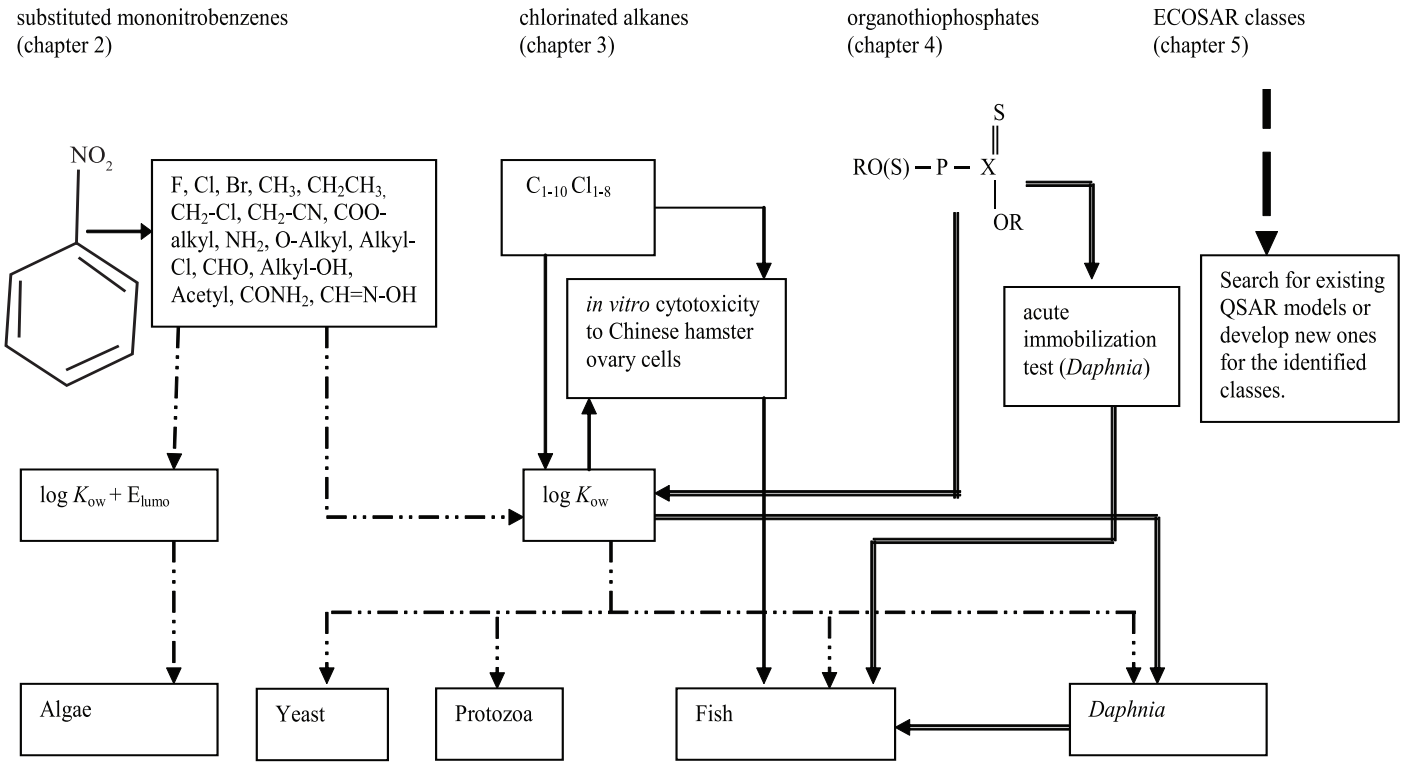


Figure 6.1: A summary of the QSAR models for acute toxicity developed in this thesis. Using estimated og K_{ow} and/or E_{lumo} values, *in vitro* tests, one can estimate toxicity of substituted mononitrobenzenes (▬ ▪ ▪), chlorinated alkanes (▬▬) and organothiophosphates (═) to various organisms.

General discussion

The main objective of this thesis was to develop validated computational chemistry-based QSAR models for selected groups of EINECS chemicals and to investigate to what extent QSAR approaches can be of use in predicting toxicty of the large number of EINECS list chemicals. This was done using experimental toxicity data from literature for substituted mononitrobenzenes (**chapter 2**) and organothiophosphates (**chapter 4**), and using experimental toxicity data from our own experimental work for chlorinated alkanes (**chapter 3**) and organothiophosphates (**chapter 4**). Since the application of the five OECD guidelines for (Q)SAR model development and validation was central in this thesis, in the following sections, the application of these guidelines is discussed in detail.

i) Clearly defined endpoint

This guideline states that it should be clear which endpoint is being modeled and how it is experimentally defined, since a given endpoint could be determined under different experimental conditions and protocols (*7*). Throughout the thesis, the toxicity endpoints were clearly defined, e.g. 24 h EC_{50} (μM) for CHO cells in the MTT cytotoxicity assay (**chapter 3**) and 24 h EC_{50} (M) for *D. magna* in the acute immobilization test (**chapter 4**) and the conditions under which the toxicity data were derived were provided. With the provision of this information, there can be no confusion as to the period of exposure, test organism, and the units of the test concentrations.

ii) Unambiguous algorithm

An unambiguous algorithm seeks to ensure transparency of the model that is used to generate predictions of a toxicity endpoint based on chemical structure and/or physicochemical properties (*7*). For each of the three descriptors used in this thesis (log K_{ow}, E_{lumo} and E_{homo}) their relevance to the toxicity endpoint was described, the calculation methods, software packages and the predicted values of the descriptors for each compound were described and the latter were generated by methods that are publicly available. For all compounds, CAS numbers were provided, to enable easier identification, as some compounds e.g. pesticides (**chapter 4**) are known by multiple names. The statistical technique, linear regression used to analyze the relationship between toxicity and the physicochemical descriptors is simple to use and sufficient especially where few descriptors (in this thesis, one or two) are involved (*17*). All experimental data used in this thesis were from the same laboratory in order to avoid inter-laboratory variation (*18*). All QSAR models developed in this thesis meet the

requirement of a maximum of one descriptor (k) for every five compounds ($n/k \geq 5$) (*19*), e.g. our QSAR model for *D. magna* (n = 10, k = 1, n/k = 10) (**chapter 4**), whilst a comparable model for the same organism and endpoints from literature (*20*) does not (*D. magna*, n = 22, k = 6, n/k = 3.6). In general, the larger the number of descriptors used, and the fewer the observations in the training set, the significantly higher the probability for the occurrence of a correlation by chance (*21*).

iii) Defined domain of applicability

As models are a simplification of reality, their limits should be well defined (*22*). This principle reflects the fact that QSARs are associated with limitations regarding the types of chemical structures, physicochemical properties, and mechanism(s) of action for which a model generates reliable predictions. The domains of applicability of our QSAR models were defined by taking into account the similarities in chemical structure and the ranges of descriptor values of the training set chemicals (**chapters 2–4**). For example, the QSAR model for describing toxicity of organothiophosphates to *D. magna* (**chapter 4**) only applies to molecules containing a P=S moiety and with log K_{ow} values lying between 0.75 and 5.50. In cases where log K_{ow} was used as a descriptor, cut-off points were set by comparing QSAR equations to those for non-polar narcotics (where available). Above the cutoff point, the non-polar narcotics QSAR should be applied in order to avoid underestimating the toxicity of compounds within the chemical domain of the QSAR (**chapter 2**) (*18*). In this thesis, we went a step further developing a methodology for extracting chemicals that fit into the applicability domain of our QSAR models (**chapter 2**), and providing lists of EINECS chemicals that fit into the applicability domains of the QSAR models for the three chemical groups (**chapters 2–4**).

iv) Appropriate measures of goodness-of-fit, robustness, and predictivity

This principle requires that parameters that reflect both the internal performance of the QSAR model as well as its predictivity should be provided (*7*). The internal performance of the models was characterized based on the goodness-of-fit (r^2) and robustness (r^2_{int}), both determined based on the training set chemicals. The robustness of the models was tested by internal validation using a leave-out-many cross validation method, with 20% of the calibration compounds left out at each step. The validation groups were created using the method of unsupervised stratification of cross validation to reduce bias introduced by random sampling. The external performance of the QSAR models was evaluated by testing a series of additional compounds that fitted into the applicability domain of the model, and then

comparing the predicted and experimental toxicity values. In this thesis, we validated our QSAR models internally (**chapters 2–4**) and externally (**chapters 3 and 4**), parts of the five OECD requirements for QSAR models that are most often neglected in most QSAR studies published in the literature.

v) A mechanistic interpretation

The OECD guidelines also state that a QSAR should be associated with a mechanistic interpretation, whenever this is possible. Such a mechanistic interpretation links the descriptors used in the model and the endpoint being predicted (*7*). When a QSAR is based on mathematical descriptors that do not relate to a mechanistic interpretation this limits the impact of a QSAR. The descriptors used to define a QSAR model should reflect the rate-limiting step in the biological process and/or endpoint that is modeled, otherwise significant correlations will not be obtained. This was demonstrated for chlorinated alkanes (**chapter 3**) when the use of E_{lumo} as an additional descriptor to log K_{ow} did not improve the correlation. Similarly, for organothiophosphates, the lack of improvement in the correlation after the use of E_{homo} or E_{lumo} as additional descriptors to log K_{ow} suggests that for *D. magna*, the conversion of the organothiophosphates to the organophosphates (modeled by E_{homo}) or the rate of reaction of the oxidized metabolite of the organothiophosphate with AChE (modeled by E_{lumo}) are not the rate-limiting steps in the mechanism underlying the toxicity (**chapter 4**).

The QSAR models developed in this thesis showed a high dependence on log K_{ow} in explaining toxicity of the three chemical groups to various organisms (**chapters 2–4**), and even to cellular systems (**chapter 3**). This agrees well with the ECOSAR software classification results (**chapter 5)** where substituted mononitrobenzenes and chlorinated alkanes were grouped into the neutral organics class, and organothiophosphates were classified into the esters class. According to the classification scheme proposed by Verhaar et al. (1992), neutral organics and esters would fall into the non-polar and polar narcotic groups respectively (*23*). It has been frequently demonstrated that the toxicity of both non-polar and polar narcotics can be adequately explained by their hydrophobicity, as they are assumed to exert their toxicity via a non-specific mode of action (*11, 23-25*), which is in agreement with the results of this thesis

When developing QSAR models one also needs to be aware that even within each specific class, a single QSAR model would probably not cover all of the compounds as shown with substituted mononitrobenzenes (**chapter 2**). The QSAR model developed for substituted mononitrobenzenes could not cover substituted nitrobenzenes that (i) are ionized

at physiological pH of 7.4, i.e. nitrophenols and nitrobenzoic acids, (ii) have two nitro groups, (iii) contain a cyano substituent, or (iv) have two phenyl rings, because they were clearly outliers in the model probably due to a different mechanism of action (**chapter 2**). Substituted dinitrobenzenes have been shown previously to have enhanced toxicity when compared to substituted mononitrobenzenes, probably due to their fast reduction to the corresponding aryl-hydroxylamines (*26-28*), which justifies their exclusion. Compounds that can ionize at physiological pH (e.g., benzoic acids and nitrophenols) have different kinetics of uptake in the body than those that are neutral and may have additional toxicity by interfering with proton gradients during oxidative phosphorylation (*18*). With respect to 2-, 3-, and 4-cyanonitrobenzene, only 2-cyanonitrobenzene was an outlier. This could be due to experimental error, or a mechanistic reason, which remains to be established.

In this thesis, $E_{lumo,}$ a reactivity parameter, was only needed as an additional descriptor to describe toxicity of mononitrobenzenes to algae (**chapter 2**), which is in agreement to an earlier trend observed after combining log K_{ow} and another reactivity parameter, Hammett's $\Sigma\sigma$ when modeling the toxicity of substituted mononitrobenzenes to *Chlorella pyrenoidosa* (*27*).

An insight into the total number of EINECS compounds that QSAR models can potentially cover would be of interest for a potential successful application of these models within REACH. Using the ECOSAR software 54 % of the EINECS compounds were grouped into classes that can be potentially covered by QSAR models, and 40 % of them fall into the non-polar and polar narcotics groups (**chapter 5**). Since log K_{ow} values can be easily estimated with software (**chapter 1**), and also since most of the QSAR models published over the last 30 years cover non-polar and polar narcotics, the potential application of QSAR models to the non-polar and polar compounds looks promising. For other compounds that ECOSAR was unable to classify, e.g. botanical extracts, analytical identification of the individual constituents using an adapted safety assessment approach used for natural flavor ingredients in food (*16*) could be a way forward Considering the rapid improvement and reduced costs of analytical tools, compared to the ever increasing costs for traditional toxicity testing, this approach may prove to be potentially useful and applicable for a wider range of botanical extracts (n = 2 819), and/or animal extracts (n = 119). This would lead to an increase in the coverage of the EINECS list by (Q)SAR approaches from 54 277 to 57 215 compounds, equivalent to 57 % of the EINECS list.

Future perspectives and conclusions

Within REACH, the range of possible uses of (Q)SARs includes screening and establishment of priorities, identification of inherent toxicity (*29*) and classification and labeling of chemical substances under the Globally Harmonized System (GHS) (*30*). The successful implementation of (Q)SAR approaches as demonstrated in this thesis further supports the belief that there is a future for (Q)SAR approaches in the risk assessment of chemicals. According to Bradbury et al. (2004), the best step forward in the risk assessment of chemicals is to use what are called Intelligent Testing Strategies (ITS) (Figure 6.2).

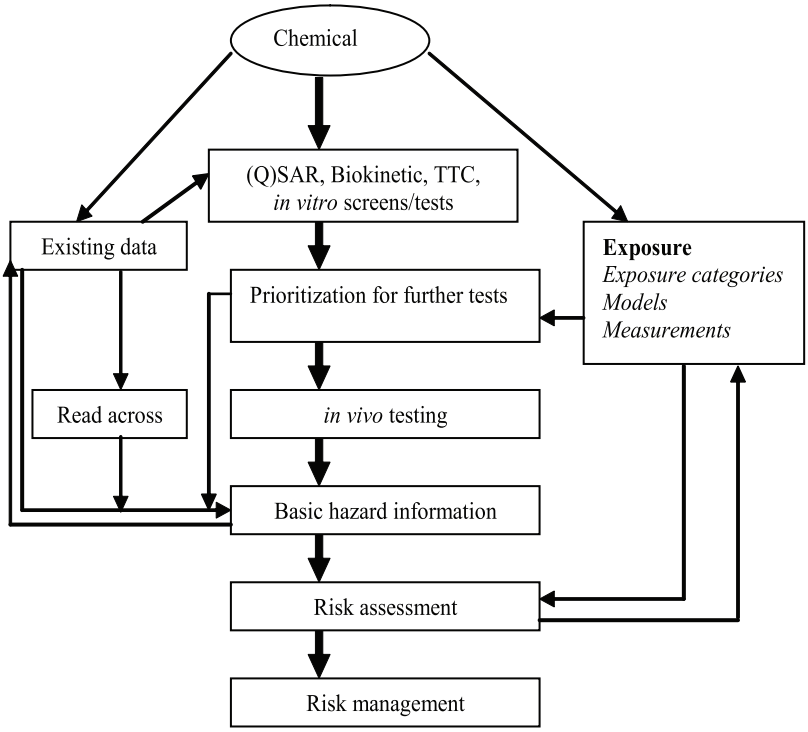


Figure 6.2: A scheme for implementing intelligent testing strategies (ITS). Implementation of QSARs into ITS can help to reduce costs, limit animal testing, and speed up the risk assessment process (*31*). TTC = thresholds of toxicological concern.

Intelligent testing strategies are defined as any approach to the evaluation of toxicity that is based on the use of two or more of the following: physicochemical data, *in vitro* data, human data (e.g. clinical case reports), animal data (where unavoidable), computational methods such as (Q)SAR and biokinetic models (*32, 33*) and thresholds of toxicological

concern (TTC) (*31*) where TTC is an exposure threshold value for chemicals, below which no significant risk is expected (*31*). The results of this thesis (Figure 6.1) can be integrated into an ITS and used to help reduce the number of animals used for experimental purposes.

The impact that (Q)SARs will have on the regulatory scene will depend on how much regulators are willing to incorporate (Q)SARs into their decision making. Estimates indicated that implementation of QSARs into the ITS would result in cost savings between € 800 and 130 million (Table 6.1) (*34*), and saving of between 1.3 and 1.9 million animals (*35*).

Table 6.1: Costs of animal tests with and without implementing QSARs into integrated testing strategies (*34*)

Test costs (million €)	Production volume (tons) per annum				Total
	1–10	10–100	100–1 000	> 1 000	
No use of QSAR	230	690	510	710	2130
Maximum use of QSAR	130	260	260	540	1190

Internationally, considerable efforts are in progress to help increase the use of (Q)SAR models for regulatory purposes. The OECD has developed a software package called (Q)SAR Application Toolbox that member countries can use to fill toxicity gaps in (eco) toxicity data needed for assessing the hazards of chemicals (*36*). The European Chemicals Bureau (ECB) is currently compiling an inventory of (Q)SAR models that can be used for regulatory purposes. For this purpose, they are developing a harmonized template, (Q)SAR Prediction Reporting Format, which will be used to summarize and report substance-specific predictions generated by (Q)SAR models (*37*). The ECB has developed a (Q)SAR tool named Toxtree to be implemented into an ITS. (*38, 39*) Toxtree is an open source application and it contains a data mining tool, Leadscope, that can be used to provide a hierarchical clustering of chemicals (*29*).

Over a period of 11 years, REACH will assemble data with a value of approximately € 10 billion (*29*). It is still unclear to what extent this large volume of data can be made available for research purposes. Public access to this data will enable the development of more (Q)SAR models, and the external validation of existing models. Currently, (Q)SAR approaches are widely used in the pharmaceutical industry to predict pharmacological properties of compounds, thus allowing for the selection of desirable candidate compounds for synthesis. This approach could also be extended to industrial chemicals in order to optimize their toxicological and/or ecotoxicological properties. Such an approach could also

be used for the development of safer chemicals as well (*29*).

Overall, it is concluded that QSAR models are expected to play an important role in the risk assessment of chemicals within REACH. The inclusion of QSAR approaches in ITS will result in considerable savings in costs and in animal numbers. The implementation of the OECD guidelines for QSAR development and validation at every step of the modeling process will help to increase the acceptability of data generated by (Q)SARs, and also to promote their mutual acceptance by regulatory authorities. Altogether, the results of this thesis reveal that, (i) *in vitro* experiments and even *in silico* calculations can help to reduce or replace animals used for experimental testing and (ii) despite the fact that individual QSARs may often each cover only limited, i.e. less than 1%, of the EINECS compounds, (Q)SAR approaches have the potential to cover about 57 % of the EINECS compounds.

References

(1) EU. (2006) Regulation (EC) No 1907/2006 of the European Parliament and of the Council. *Off. J. Eur. Union*, 1–849.

(2) EU. (1993) Council Regulation (EEC) No 793/93 of 23 March 1993 on the evaluation and control of the risks of existing substances *Off. J. Eur. Union. L 084*, 1–75.

(3) EU. (2007) Fifth report on the statistics on the number of animals used for experimental and other scientific purposes in the member states of the European Union, pp 1-14, Brussels.

(4) Schultz, T. W., Cronin, M. T. D., Walker, J. D. and Aptula, A. O. (2003) Quantitative structure-activity relationships (QSARs) in toxicology: a historical perspective. *Journal of Molecular Structure: THEOCHEM 622*, 1-22.

(5) Abraham, M. H., Chadha, H. S., Martins, F., Mitchell, R. C., Bradbury, M. W. and Gratton, J. A. (1999) Hydrogen bonding part 46: A review of the correlation and prediction of transport properties by an LFER method: Physicochemical properties, brain penetration and skin permeability. *Pesticide Science 55*, 78.

(6) Hansch, C. and Fujita, T. (1964) A method for the correlation of biological activity and chemical structure. *J. Amer. Chem. Soc. 86*, 1616 - 1626.

(7) OECD. (2007) Guidance document on the validation of (Quantitative) structure-activity relationships [(Q)SAR] models, ENV/JM/MONO(2007)2, In *OECD Environment Health and Safety Publications: Series on Testing and Assessment No. 69. Paris, France* pp 1–154.

(8) Bohac, M., Loeprecht, B., Damborsky, J. and Schüürmann, G. (2002) Impact of orthogonal signal correction (OSC) on the predictive ability of CoMFA models for the ciliate toxicity of nitrobenzenes. *Quant. Struct. Act. Relat. 21*, 3 - 11.

(9) Wang, X., Yin, C. and Wang, L. (2002) Structure-activity relationships and response-surface analysis of nitroaromatics toxicity to the yeast (Saccharomyces cerevisiae). *Chemosphere 46*, 1045-1051.

(10) Trohalaki, S., Gifford, E. and Pachter, R. (2000) Improved QSARs for predictive toxicology of halogenated hydrocarbons. *Comput. Chem. 24*, 421-427.

(11) Könemann, H. (1981) Quantitative structure-activity relationships in fish toxicity studies Part 1: Relationship for 50 industrial pollutants. *Toxicology 19*, 209–221.

(12) EPA. (1999) Recognition and Management of Pesticide Poisonings, (Reigart, J. R. and Roberts, J. R., Eds.) p 243.

(13) Vittozzi, L., Fabrizi, L., Di Consiglio, E. and Testai, E. (2001) Mechanistic aspects of organophosphorothionate toxicity in fish and humans. *Environ. Int. 26*, 125–129.

(14) Nigg, H. N. and Knaak, J. B. (2000) Blood cholinesterases as human biomarkers of organophosphorous pesticide exposure. *Rev. Environ. Contam. Toxicol. 163*, 29–111.

(15) Jokanovic, M. (2001) Biotransformation of organophosphorous compounds. *Toxicology 166*, 139-160.

(16) Smith, R. L., Cohen, S. M., Doull, J., Feron, V. J., Goodman, J. I., Marnett, L. J., Portoghese, P. S., Waddell, W. J., Wagner, B. M., Hall, R. L., Higley, N. A., Lucas-Gavin, C. and Adams, T. B. (2005) A procedure for the safety evaluation of natural flavor complexes used as ingredients in food: Essential oils. *Food Chem. Toxicol. 43*, 345–363.

(17) Petrauskas, A. A. and Kolovanov, E. A. (2000) ACD/Log P method description. *Perspect. Drug Discovery Des. 19*, 99-116.

(18) Cronin, M. T. D. and Schultz, T. W. (2003) Pitfalls in QSAR. *Journal of Molecular Structure: THEOCHEM 622*, 39-51.

(19) Tropsha, A., Gramatica, P. and Gombar, V. K. (2003) The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR*

and Combinatorial Science 22, 69-77.
(20) Vighi, M., Garlanda, M. M. and Calamari, D. (1991) QSARs for toxicity of organophosphorous pesticides to *Daphnia* and honeybees. *Sci. Total Environ. 109-110*, 605-622.
(21) Drew, G. B., Lumley, J. A. and Price, N. R. (1999) Predicting ecotoxicology of organophosphate insecticides: successful parameter selection with the genetic function algorithm. *Quant. Struct. Act. Relat. 18*, 573-583.
(22) Worth, A. P. and Cronin, M. T. D. (2004) Report of the workshop on the validation of QSARs and other computational prediction models. *ATLA Alternatives to Laboratory Animals 32*, 703-706.
(23) Verhaar, H. J. M., van Leeuwen, C. J. and Hermens, J. L. M. (1992) Classifying environmental pollutants. *Chemosphere 25*, 471-491.
(24) Veith, G. D. and Broderius, S. J. (1990) Rules for distinguishing toxicants that cause type I and type II narcosis syndromes. *Environ. Health Persp. 87*, 207–211.
(25) Veith, G. D., Call, D. J. and Brooke, L. T. (1983) Structure - toxicity relationships for the fathead minnow, *Pimephales promelas*: narcotic industrial chemicals. *Can. J. Fish. Aquat.Sci. 40*, 743–748.
(26) Deneer, J. W., Sinnige, T. L., Seinen, W. and Hermens, J. L. M. (1987) Quantitative structure-activity relationships for the toxicity and bioconcentration factor of nitrobenzene derivatives towards the guppy (Poecilia reticulata). *Aquat. Toxicol. 10*, 115-129.
(27) Deneer, J. W., van Leeuwen, C. J., Seinen, W., Maas-Diepeveen, J. L. and Hermens, J. L. M. (1989) QSAR study of the toxicity of nitrobenzene derivatives towards *Daphnia magna, Chlorella pyrenoidosa* and *Photobacterium phosphoreum*. *Aquat. Toxicol. 15*, 83-98.
(28) Cronin, M. T., Gregory, B. W. and Schultz, T. W. (1998) Quantitative structure-activity analyses of nitrobenzene toxicity to *Tetrahymena pyriformis*. *Chem. Res. Toxicol. 11*, 902-908.
(29) Lahl, U. and Gundert-Remy, U. (2008) The use of (Q)SAR methods in the context of REACH. *Toxicology Mechanisms and Methods 18*, 149-158.
(30) Winder, C., Azzi, R. and Wagner, D. (2005) The development of the globally harmonized system (GHS) of classification and labelling of hazardous chemicals. *Journal of Hazardous Materials 125*, 29-44.
(31) Bradbury, S. P., Feijtel, T. C. J. and Van Leeuwen, C. J. (2004) Meeting the scientific needs of ecological risk assesstment in a regulatory context. *Environmental Science and Technology 38*.
(32) Blaauboer, B. J. (2002) The necessity of biokinetic information in the interpretation of in vitro toxicity data. *ATLA Alternatives to Laboratory Animals 30*, 85-91.
(33) Combes, R., Barratt, M. and Balls, M. (2003) An overall strategy for the testing of chemicals for human hazard and risk assessment under the EU REACH system. *ATLA Alternatives to Laboratory Animals 31*, 7-19.
(34) Pedersen, F., de Bruijn, J., Munn, S., Worth, A. P. and van Leeuwen, K. (2003) The cost saving potential of QSARs, In *Stakeholder workshop on impact assessment of REACH*, Joint Research Centre, European Chemicals Bureau, European Commission, Directorate General, Brussels.
(35) Van der Jagt, K., Munn, S., Torslov, J. and De Brujin, J. (2004) Alternative approaches can reduce the use of test animals under REACH, In *Addendum to the Report "Assessment of Additional Testing Needs under REACH. Effects of (Q)SARs, Risk Based Testing and Voluntary Industry Initiatives"*.
(36) OECD. OECD (Q)SAR application toolbox; (http://www.oecd.org/document/23/0,3343,en_2649_34379_33957015_1_1_1_1,00.html).
(37) ECB. European Chemicals Bureau, (Q)SAR Prediction Reporting Format; (http://ecb.jrc.it/qsar/qsar-tools/index.php?c=QRF).
(38) Vermeire, T. G., Aldenberg, T., Dang, Z., Janer, G., de Knecht, J. A., van Loveren, H.,

Peijnenburg, W. J. G. M., Piersma, A. H., Traas, T. P., Verschoor, A. J., van Zilverden, M. and Hakkert, B. (2007) Selected integrated testing strategies (ITS) for the risk assessment of chemicals.

(39) ECB. (2007) Toxtree. Presentation to 4th (Q)SAR working group meeting, 28 June 2007, European Chemicals Bureau, Institute for Health and Consumer Protection, JRC, Ispra, Italy.

Samenvatting

Het is algemeen bekend dat er onvoldoende (eco)toxicologische informatie beschikbaar is over de mogelijk gevaarlijke eigenschappen van meer dan 100 000 chemicaliën die binnen de Europese Unie geregistreerd zijn in de "European Inventory of Existing Chemicals" (EINECS). Met het onlangs geïntroduceerde nieuwe EU-managementinformatiesysteem voor chemische stoffen, genaamd Registration, Evaluation and Authorization of Chemicals (REACH), zouden in de komende tien jaar de nog ontbrekende toxiciteitgegevens van de EINECS-stoffen moeten worden aangevuld.

Er is momenteel een discussie gaande over het mogelijk grote aantal proefdieren dat nodig zou zijn voor experimentele toxiciteitstesten als gevolg van de invoering van REACH. Deze bezorgdheid komt mede voort uit het feit dat er binnen de EU op dit moment 12 miljoen proefdieren worden gebruikt voor experimentele doeleinden. Methoden die kunnen helpen om prioriteiten te stellen voor toekomstige toxiciteitstesten, kunnen tijd, geld en proefdieren besparen. In dit verband bestaat de verwachting dat *in silico* methoden, gebaseerd op (Quantitative) Structure Activity Relationships [(Q)SARs], informatie over toxiciteit of risico's van chemicaliën kunnen verschaffen. Deze informatie kan vervolgens worden gebruikt om prioriteiten te stellen bij het bepalen welke chemicaliën voor verdere risico-evaluatie in aanmerking komen.

Een QSAR is gebaseerd op de aanname dat de biologische activiteit van een nieuwe- of nog niet geteste chemische stof kan worden afgeleid uit de moleculaire structuur, of uit de eigenschappen van vergelijkbare stoffen waarvan de werking al is beoordeeld. Met de chemische structuur van een stof als enig gegeven, kan een voorspellend model voor toxiciteit worden ontwikkeld gebaseerd op parameters die de fysisch-chemische eigenschappen en de relatieve reactiviteit van reeds geteste stoffen definiëren.

Dit proefschrift beschrijft de ontwikkeling van gevalideerde QSAR modellen voor de drie volgende chemische groepen: mononitrobenzenen (**hoofdstuk 2**), gechlorineerde alkanen (**hoofdstuk 3**) en organothiofosfaten (**hoofdstuk 4**), rekening houdend met de richtlijnen die onlangs zijn goedgekeurd door de Organization for Economic Cooperation and Development (OECD). Deze chemische groepen worden op grote schaal gebruikt en impliceren dientengevolge een groot potentieel risico op (eco)toxicologische effecten.
Twee relevante fysisch-chemische parameters bij de modellering van de toxiciteit van de chemische stoffen uit de drie groepen voor verscheidene organismen waren:

(i) de logaritme van de octanol/water partitiecoëfficiënt (log K_{ow}), die doorgaans de hydrofobe eigenschap van een stof weergeeft, van belang bij de beschrijving van het passeren van een stof door membranen en

(ii) de energie van de laagste onbezette moleculaire orbitaal E_{lumo}(eV), die de electrofiele aard van een stof weergeeft.

Gebaseerd op experimentele toxiciteitgegevens uit de literatuur en *in silico* log K_{ow} waarden, werd een QSAR adviesinstrument ontwikkeld dat gebruikers op het spoor zet van het meest geschikte QSAR model dat toegepast moet worden om de toxiciteit te voorspellen van gesubstitueerde mononitrobenzenen voor vijf typen organismen (protozoa, vis, daphnia's, algen en gist) binnen een gespecificeerde range aan log K_{ow} –waarden (**hoofdstuk 2**). E_{lumo} was vereist als een extra parameter naast log K_{ow} voor de toxiciteitvoorspellingen voor algen.

In een volgende studie werden QSAR modellen ontwikkeld om *in vivo* acute toxiciteit te voorspellen van gechloreerde alkanen voor vis, gebaseerd op data van eigen *in vitro* experimenten, en zelfs op basis van uitsluitend *in silico* log K_{ow} data (**hoofdstuk 3**). Bovendien is een interspecies QSAR model ontwikkeld om toxiciteit van organothiofosfaten voor vis te voorspellen met gebruikmaking van toxiciteitgegevens uit proeven met acute inactivering van *Daphnia magna*, gebaseerd op *D. magna* data en *in silico* log K_{ow-}waarden (**hoofdstuk 4**).

De QSAR-modellen voor de mononitrobenzenen (hoofdstuk 2), gechloreerde alkanen (hoofdstuk 3), en organothiofosfaten (hoofdstuk 4) bleken respectievelijk 0.5 %, 0.1 %, en 0.1 % te beslaan van de 100 196 EINECS stoffen. tezamen oplopend tot 660 stoffen ofwel 0.7 % van de EINECS lijst. Gezien het feit dat de QSAR modellen slechts een klein percentage van de EINECS lijst bleken te omvatten, is in een volgende stap een schatting gemaakt van het percentage van EINECS-chemicaliën die in specifieke chemische klassen kunnen worden ingedeeld en daardoor in theorie geschikt zouden kunnen zijn voor QSAR modellering (**hoofdstuk 5**).

ECOSAR, een gratis te downloaden software pakket, verdeelde 54% van de EINECS-stoffen in 49 klassen die in beginsel kunnen worden gebruikt voor QSAR modellen. Voor stoffen die niet geclassificeerd konden worden, zoals botanische extracten, zou de risicoschatting kunnen worden uitgevoerd door verdere ontwikkeling van een onlangs beschreven methode voor de risicoschatting van natuurlijke smaakstoffen die als voedselingrediënten worden gebruikt. Dit zou resulteren in een extra 3% van de EINECS-stoffen die in principe kunnen worden bestudeerd met SAR-methoden, waarmee het totale percentage van EINECS-stoffen die met (Q)SAR benaderingen kunnen worden beoordeeld op 57% uitkomt.

Concluderend kan worden gesteld dat de resultaten van dit proefschrift uitwijzen dat

(i) *in vitro* experimenten en zelfs *in silico* berekeningen kunnen helpen om dierproeven voor experimentele toxiciteitproeven te reduceren of te vervangen

en dat

(ii) ondanks het feit dat individuele QSARs vaak slechts een beperkte hoeveelheid EINECS-stoffen omvatten, d.w.z. minder dan 1% daarvan, (Q)SAR methodes het mogelijk maken ongeveer 57% van de EINECS-stoffen te beoordelen.

Appendix

Supplementary information for Chapter 4

Table S4.1: Chemical Abstract Service (CAS) numbers, calculated descriptors and experimental toxicity data (Johnson and Finley, 1980) describing 96 h toxicity of organothiophosphates to fathead minnow (*Pimephales promelas*)

No.	CAS #	name	log K_{ow}	E_{homo} (P=S) eV	E_{lumo} (P=O) eV	96 h log LC_{50} (M)
1	86-50-0	Azinphos methyl	2,549	-9,415	1,551	-4,262
2	298-04-4	Disulfoton	4,173	-9,247	4,37	-5,634
3	563-12-2	Ethion	5,569	-9,51	3,473	-5,842
4	122-14-5	Fenitrothion	3,209	-9,758	1,642	-5,875
5	55-38-9	Fenthion	3,839	-8,921	3,116	-5,581
6	121-75-5	Malathion	2,475	-9,459	3,715	-4,364
7	56-38-2	Parathion ethyl	3,468	-9,879	1,203	-5,784
8	298-00-0	Parathion-methyl	2,79	-9,931	1,188	-5,619
9	2310-17-0	Phosalone	4,31	-9,052	2,799	-5,671
10	732-11-6	Phosmet	3,109	-9,352	1,255	-4,614
11	14816-18-3	Phoxim	4,385	-9,488	1,191	-4,793
12	3383-96-8	Temephos	5,736	-8,355	3,148	-5,01

Table S4.2: Chemical Abstract Service (CAS) numbers, calculated descriptors and experimental toxicity data describing 96 h toxicity of organothiophosphates to rainbow trout (*Oncorhynchus mykiss*)

No.	CAS #	name	E_{homo} (P=S) eV	E_{lumo} (P=O) eV	log LC_{50} (M)[a]	log LC_{50} (M)[b]
1	2642-71-9	Azinphos-ethyl	-9,29	1,53	-7,24	-7,24
2	86-50-0	Azinphos methyl	-9,42	1,55		-7,87
3	2921-88-2	Chlorpyrifos ethyl	-9,58	1,60	-7,39	-7,69
4	5598-13-0	Chlorpyriphos methyl	-9,62	1,65	-6,44	
5	56-72-4	Coumaphos	-9,02	1,48	-5,50	-5,61
6	333-41-5	Diazinon	-9,23	2,98	-5,68	
7	97-17-6	Dichlofenthion	-9,29	2,88		-5,40
8	60-51-5	Dimethoate	-9,45	3,92	-4,53	-7,57
9	298-04-4	Disulfoton	-9,25	4,37		-5,17
10	563-12-2	Ethion	-9,51	3,47		-5,89
11	299-84-3	Fenchlorphos	-9,48	2,46		-5,77
12	122-14-5	Fenitrothion	-9,76	1,64		-8,06
13	55-38-9	Fenthion	-8,92	3,12	-5,52	-5,48
14	18181-70-9	Iodofenphos	-9,10	2,54		-7,41
15	121-75-5	Malathion	-9,46	3,72	-6,47	-6,22
16	950-37-8	Methidathion	-9,38	3,75	-7,33	-7,33
17	56-38-2	Parathion ethyl	-9,89	1,20		-5,31
18	298-00-0	Parathion-methyl	-9,93	1,19	-4,91	-4,85
19	298-02-2	Phorate	-9,26	4,15		-7,30
20	2310-17-0	Phosalone	-9,05	2,80	-6,39	-6,39
21	732-11-6	Phosmet	-9,35	1,26	-5,78	-6,02
22	14816-18-3	Phoxim	-9,49	1,19		-6,23
23	29232-93-7	Pirimiphos-methyl	-8,55	3,30	-5,65	
24	3383-96-8	Temephos	-8,36	3,15		-8,13

LC_{50} values from [a] Bermudez-Saldana et al, 2004; [b] Johnson and Finley, 1980

Table S4.3: Chemical Abstract Service (CAS) numbers, calculated descriptors and experimental toxicity data describing 96 h toxicity of organothiophosphates to bluegill (*Lepomis macrochirus*) EC_{50} values from [a] Johnson and Finley, 1980; [b] Bermudez-Saldana et al, 2004

No.	CAS #	name	log K_{ow}	E_{homo} (P=S) eV	E_{lumo} (P=O) eV	log LC_{50} (M)[a]	log LC_{50} (M)[b]
1	2642-71-9	Azinphos-ethyl	3,43	-9,29	1,53	-6,49	-8,50
2	86-50-0	Azinphos methyl	2,55	-9,42	1,55	-5,06	
3	786-19-6	Carbophenothion	5,94	-9,18	1,81	-5,66	
4	2921-88-2	Chlorpyrifos ethyl	4,51	-9,58	1,60	-6,27	-8,07
5	5598-13-0	Chlorpyriphos methyl	3,83	-9,62	1,65		-5,56
6	56-72-4	Coumaphos	4,33	-9,02	1,48	-4,11	-6,30
7	333-41-5	Diazinon	3,32	-9,23	2,98	-4,32	-6,25
8	97-17-6	Dichlofenthion	5,04	-9,29	2,88	-3,61	
9	60-51-5	Dimethoate	0,75	-9,45	3,92	-5,10	-4,58
10	298-04-4	Disulfoton	4,17	-9,25	4,37	-4,14	
11	563-12-2	Ethion	5,57	-9,51	3,47	-4,42	
12	299-84-3	Fenchlorphos	4,97	-9,48	2,46	-3,58	
13	122-14-5	Fenitrothion	3,21	-9,76	1,64	-5,93	
14	55-38-9	Fenthion	3,84	-8,92	3,16	-3,44	-5,26
15	121-75-5	Malathion	2,48	-9,46	3,72	-4,38	-6,57
16	950-37-8	Methidathion	2,78	-9,38	3,75	-5,49	-7,53
17	56-38-2	Parathion ethyl	3,47	-9,88	1,20	-3,94	
18	298-00-0	Parathion-methyl	2,79	-9,93	1,19	-2,80	-4,91
19	298-02-2	Phorate	3,84	-9,26	4,19	-6,28	
20	2310-17-0	Phosalone	4,31	-9,05	2,80	-4,63	-6,57
21	732-11-6	Phosmet	3,11	-9,35	1,26	-4,19	-5,88
22	14816-18-3	Phoxim	4,39	-9,49	1,19	-4,73	
23	29232-93-7	Pirimiphos-methyl	3,38	-8,55	3,30		-5,03
24	3383-96-8	Temephos	5,74	-8,36	3,15	-5,42	

ABBREVIATIONS

Abbreviations

General

AChE	Acetylcholinesterase
CHO	Chinese hamster ovary
DMSO	Dimethyl sulfoxide
ECB	European Chemicals Bureau
ECOSAR	Ecological Structure Activity Relationships
EINECS	European inventory of existing commercial chemical substances
E_{homo}	Energy of the highest occupied molecular orbital
E_{lumo}	Energy of the lowest unoccupied molecular orbital
HPV	High production volume
ITS	Intelligent testing strategy
K_{ow}	Octanol-water partition coefficient
LCA	Lower chlorinated alkane
LMO	Leave-many-out
LOO	Leave-one-out
MMFF	Merck molecular force field
MTT	[3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide]
OECD	Organization for Economic Cooperation and Development
PCA	Polychlorinated alkane
P=O	Organophosphate
P=S	Organothiophosphate
(Q)SAR	(Quantitative) structure-activity relationship
3Rs	Replacement, reduction and refinement
REACH	Registration Evaluation and Authorization of Chemicals
SMILES	Simplified molecular input line entry system
SDR	Studentized deleted residual
SPSS	Statistical Package for Social Scientists
ZonMw	Netherlands Organization for Health Research and Development Nederlandse Organisatie voor Gezondheidsonderzoek en Zorginnovatie

Statistical

F	variance ratio
n	number of compounds used for developing model
PRESS	predictive sum of squares
r^2	coefficient of determination
r^2_{int}	internally cross validated coefficient of determination
r^2_{ext}	externally validated coefficient of determination
s	standard error of the estimate
SSD	sum-of-squares deviation

ACKNOWLEDGEMENTS

Acknowledgements

Indeed, there is a time for everything, a time to sow and a time to reap. Thanks be to our God and Father of all, who is over all, through all, and in all. After four very long and arduous years, I am delighted to add on one more academic title. Of course, I did not achieve all this alone. My appreciation goes to the following:

QSAR Project members: *Ivonne Rietjens* – for the inspiration, insight, quick and constructive feedback, and teaching me to work more efficiently. It was a pleasure working with you. After every meeting, I always came out more focused and determined to achieve set goals. Thank you for helping me to become a better and critical scientist, and for bringing the best out of me. *Tinka Murk*– for being my supervisor and mentor for the last 6 years. Your critical comments, support, encouragement helped me to become a better and focused researcher. Thank you for your constant emphasis and help in disseminating our research to outside our laboratory. *Andreas Freidig* – even though you worked away from Wageningen, you always came up with wonderful and timely ideas that made the manuscripts stand out. Thank you for the inspiring ideas and literature that you provided.
Jacques Vervoort – for arranging, setting up and troubleshooting all the computer software and hardware. When those E_{lumo} calculations seemed impossible at the start, you were there to see them through. *Ans Soffers* – for help with the literature searches and computer calculations, and especially for introducing me to the Spartan software. I could not have performed all those thousands of calculations on my own. *Hans vanden Berg* – for expert and timely assistance with the experimental work. Thanks for accompanying me to NOTOX and helping with the experiments, even when it seemed the experiments would never end. When l was busy with conferences and courses, you kept my research moving.
Toxers - Thank you all for the good atmosphere, hospitality, and for providing a second "home away from home". My appreciation goes to *Gre, Irene, Jac, Gerrit,* Bert, *Laura, Marelle*, fellow PhD students (*Ana, Ans, Walter, Ashwin, Henrique, Hester, Merijn, Marcel, Wiratno, Jaime, Vincent dB, Vincent vB, Maaike, Suzanne, Yee, Jochem, Niek*) for your company, the laughs, the BBQs, the lab trips, Sinterklaas, Xmas dinners. Special thanks to my office mates, *Ans and Walter*, for providing a "*gezellig sfeer*". Secretariat – *Gre and Irene*- thank you for handling all the administrative matters that made my work and stay in the Netherlands comfortable. *Irene* – thank you for the Dutch translations in this thesis. Thanks also to *Gerrit* and *Jac* for your input during the scientific discussions and to *Bert* for your invaluable help with information on chemicals. *Cozmina* (TOX-WUR and TOX-

IRAS) – for your friendship and support over the last 6 years. Who ever thought we would be colleagues again at IRAS?
-Thesis students: *Marije, Oman, Tamas, Tingting, Tom* – thank you for your invaluable scientific input into this thesis. It was a pleasure supervising such hardworking and determined students.
NOTOX BV – Thank you for providing the facilities to perform experiments with *Daphnia,* and technical expertise of *Erik and Vicky.*

Of course, I did not spend all my time in the lab. I would also like to express my thanks to:

-My wife, *Ednah,* and our daughter, *Heather*, for their love, support, encouragement and patience. It has been a long road and I could not have walked it alone. Thanks Ednah for proofreading this thesis. -Our dearly beloved friends – *Maphosas, Matongos, and Mugovas*– for their unwavering friendship and support. -*Fellow Zimbabweans over the years* – for the dinners, sharing of stories from home–you really helped to keep me connected to our beloved country. *Tinotenda, Siyabonga, ngazvirambe zvakadaro, hukama igasva hunozadziswa nekudya.*
-*Amazing Grace Parish* – for prayers, fellowship, support, guidance. *Amazing Football team* – for the environment you provided that allowed me to unwind after a hard day's work. Keep up the spirit, and one day you will win the national championship!
-*Bert and Mieke* –for making my family feel welcome in the Netherlands over the last 6 years. We really enjoy(ed) your Dutch cuisine and hospitality.
-*Lombardi "Woestijn" neighbors* – for providing a hospitable community. I really enjoy(ed) our yearly BBQs and cannot wait for the next one.
-*University of Zimbabwe, Biochem Dept* – Dr *B. Masola* & other scientific & technical staff for providing a solid background on which to launch my scientific career.

Finally, I would like to thank my p*arents and siblings* – Even though you were far away, you were always in my thoughts. Thank you for the investment in my education and for supporting me in all the endeavors l undertook.

About the Author

Elton Zvinavashe was born on March 30, 1975 in Harare, Zimbabwe. He completed his high school education at Kutama College (Norton, Zimbabwe) in 1993. From 1994, he studied Biochemistry at the University of Zimbabwe (UZ) and graduated with B. Sc. Honors in 1996. From 1997 to 2002, he worked as a Teaching and Research Assistant in the Dept of Biochemistry (UZ). During this period, he worked on several research projects including regulation of amino acid metabolism, protein purification, enzyme kinetics, enhancement of flavor and shelf life of local beverages. During the same period (1997-2002), he also served first as the Treasurer and then later as the Secretary of the Biochemistry and Molecular Biology Society of Zimbabwe, an associate adhering body to the International Union of Biochemistry and Molecular Biology. From September 2002 to January 2004, he studied MSc in Environmental Sciences (specializing in Environmental Toxicology) at Wageningen University, NL, and completed his thesis entitled "Development of an *in vitro* amphibian test system using biochemical markers as indicators of thyroid hormone disruption" under the supervision of Prof. A.J. Murk (Dept of Toxicology), for which he graduated *cum laude*. Between May 2004 and June 2008, he was appointed as a PhD student at Wageningen University (Dept of Toxicology), where he worked on the project entitled "The potential of computer-based quantitative structure activity approaches for predicting acute toxicity of chemicals" as described in this thesis. During his PhD period, he also attended several conferences and courses, including courses organized under the auspices of the Netherlands Postdoctoral Education in Toxicology (PET), enabling him to be registered as a Toxicologist.

Since June 2008, he is working as a post-doctoral researcher at the Institute for Risk Assessment Sciences (IRAS-TOX) at the University of Utrecht, NL.

List of publications

Peer reviewed articles

Zvinavashe E, Berg H. van den, Soffers AEMF, Vervoort J, Freidig AP, Murk AJ, Rietjens IMCM. **2008**. QSAR models for predicting *in vivo* aquatic toxicity of chlorinated alkanes to fish. *Chemical Research in Toxicology* 21:739─745.

Zvinavashe E, Murk AJ, Soffers A, Vervoort J, Freidig AP, Rietjens IMCM. **2006**. Quantum chemistry-based quantitative structure-activity relationships for modeling the (sub)acute toxicity of substituted mononitrobenzenes in aquatic systems. *Environmental Toxicology and Chemistry* 25: 2313─2321.

Zvinavashe E, Murk AJ, Rietjens IMCM. The promises and pitfalls of QSAR approaches for predicting metabolism and toxicity. *Chemical Research in Toxicology, in press*.

Zvinavashe E, Murk AJ, Rietjens IMCM. On the number of EINECS compounds that can be covered by (Q)SAR models for acute toxicity. *Toxicology Letters, in press*.

Zvinavashe E, Du T, Griff T, van den Berg HJ, Soffers AEMF, Vervoort J, Murk AJ, Rietjens IMCM. Quantitative structure-activity relationship modeling of the toxicity of organothiophosphate pesticides to aquatic organisms and honeybee. *Chemosphere* (*submitted*).

Freidig AP, Dekkers S, Verwei M, Zvinavashe E, Bessems JGM, van de Sandt JJM. **2007** Development of a QSAR for worst case estimates of acute toxicity of chemically reactive compounds. *Toxicology Letters*, 170: 214-222.

Schriks M, Zvinavashe E, Furlow, DT, Murk AJ. **2006**. Disruption of thyroid hormone-mediated *Xenopus laevis* tadpole tail tip regression by hexabromocyclododecane (HBCD) and 2,2′,3,3′,4,4′,5,5′,6-nona brominated diphenyl ether (BDE206). *Chemosphere* 65: *1904─1908*.

Masola B, Zvinavashe E. **2003**. Phosphate-dependent glutaminase in enterocyte mitochondria and its regulation by ammonium and other ions. *Amino Acids* 24: 427─*434*.

Abstract publications

Zvinavashe E, Berg van den, Hans, Soffers A, Vervoort J, Freidig AP, Murk AJ, Rietjens, IMCM. **2007.** Possibilities and limits of QSAR models in ecotoxicology, *IMARES modeling day.* September 12, IMARES, IJmuiden, NL.

M. Strikwold and E. Zvinavashe **2007.** Improvement and development of QSARs for phenols. *PhD Symposium, Netherlands Society of Toxicologists (NVT)*, June 12–13, 2007, Wageningen, NL. Published in *Chemico-Biological Interactions*, 2007, *69(2): 14.*

Zvinavashe E, Berg van den, Hans, Soffers A, Vervoort J, Freidig AP, Murk AJ, Rietjens, IMCM. **2007.** How many EINECS list compounds can be covered by QSAR models? *PhD Symposium, (NVT)*, June 12–13, Wageningen, NL. Published in *Chemico-Biological Interactions*, 2007, *69(2): 143.*

Zvinavashe E, Berg van den, Hans, Soffers A, Vervoort J, Freidig AP, Murk AJ, Rietjens, IMCM. **2007.** How many EINECS list compounds can be covered by QSAR models? *Proceedings of the SETAC Europe 17th Annual Meeting "The multiple stressors for the environment - present and future challenges and perspectives"*, May 20–24, Porto, Portugal.

Zvinavashe E, Soffers A, Vervoort J, Freidig AP, Murk AJ, Rietjens, IMCM. **2006.** Multi-species tool for modeling the acute toxicity of substituted mononitrobenzenes in aquatic systems. *Proceedings of the 12th International Workshop on Quantitative Structure-Activity Relationships in Environmental Toxicology: Bridging QSAR and Environmental Fate Modeling*, *SAR and QSAR in Environmental Research,* May 8–12, Lyon, France

Zvinavashe E, Murk, AJ, Vervoort J., Soffers AEMF, Freidig AP, Rietjens IMCM. **2006.** Multi-species tool for modeling acute toxicity of substituted mononitrobenzenes in aquatic systems. *PhD Symposium, Netherlands Society of Toxicologists (NVT)*, June 13–14, Wageningen, NL. Published in *Chemico-Biological Interactions* 2006, 161: 175.

Zvinavashe E, Soffers A, Vervoort J, Freidig AP, Murk AJ, Rietjens, IMCM. **2006.** Multi-species tool for modeling the acute toxicity of substituted mononitrobenzenes in aquatic systems *Proceedings of the 1 day symposium organized by RIVM, KNCV and NVT "Does Science REACH policy"* December 1, RIVM, Bilthoven, The Netherlands:

Murk AJ, Zvinavashe E, Soffers A, Vervoort J, Freidig AP, Rietjens, IMCM. **2006.** Multi-species tool for modeling the acute toxicity of substituted mononitrobenzenes in aquatic systems. *Proceeding of the SETAC Europe, 16th Annual Meeting: Controversies and Solutions in Environmental Sciences*, May 7─12, The Hague, NL.

Zvinavashe E, Murk AJ, Soffers A, Vervoort J, Freidig A, Rietjens, IMCM. **2005.** Quantum chemistry based QSARs for modeling the (sub) acute toxicity of EINECS list nitrobenzenes. *PhD Symposium, Netherlands Society of Toxicologists (NVT)*, June 7─8, Organon, Oss, NL.

Newsletters

Zvinavashe E, Murk AJ, Soffers A, Vervoort J, Freidig A, Rietjens, IMCM. **2007**. Development of a computational chemistry-based QSAR approach that sets priorities in toxicity testing of selected groups of EINECS list chemicals. Netherlands Centre for Alternatives to animal use (NCA), Issue 22 – March 2007 p 3─5. http://www.nca-nl.org/

Training and supervision plan (approved by SENSE Research School)

Overview of courses and meetings attended during PhD period

Courses

Essay: Promoting the use of alternatives in toxicity testing of animals (SENSE)	2008
Introduction to Laboratory Animal Sciences (PET)	2007
Medical, Forensic & Regulatory Toxicology (PET)	2007
Molecular Toxicology (PET)	2007
Mutagenesis and Carcinogenesis (PET)	2007
Elaboration of PhD Proposal and Time Planning (SENSE)	2006
Risk Assessment (PET, SENSE)	2006
Basic Statistics (SENSE)	2006
Organ Toxicology (PET)	2005
Immunotoxicology (PET)	2005
Environmental Research in Context (SENSE)	2005
Techniques for Writing and Presenting Scientific Papers (SENSE)	2005
Special Topics in Ecotoxicology (SENSE)	2004
Pathobiology (PET)	2004
Teaching and Supervising Thesis Students (WGS)	2004

Meetings

IMARES Modeling Day, Ijmuiden, The Netherlands	2007
SETAC Europe, 17th Annual Meeting, Porto, Portugal	2007
PhD Symposium, (NVT), Wageningen, The Netherlands	2007
SETAC Europe, 16th Annual Meeting, The Hague, The Netherlands	2006
Int'l Workshop (12th) on QSAR in Environmental Toxicology, Lyon, France	2006
PhD Symposium, (NVT), Wageningen, The Netherlands	2006
SENSE Summer Symposium, Env Change: Prevent or Adapt, Amsterdam, NL	2006
PhD Symposium, (NVT), Organon, Oss, The Netherlands	2005
PhD Symposium, (NVT), Veldhoven, The Netherlands	2004

Abbreviations

IMARES	Institute of Marine Resources and Ecosystem Studies
NVT	Netherlands Society for Toxicologists
PET	Postgraduate Education in Toxicology
QSAR	Quantitative Structure-Activity Relationship
SENSE	Socio-Economic and Natural Sciences of the Environment
WGS	Wageningen Graduate Schools

Research presented in this PhD dissertation was carried out at the Department of Toxicology, Wageningen University, The Netherlands, and was financially supported by the Netherlands Organization for Health Research and Development (ZonMW; project no. 3170.0066), a division of the Netherlands Scientific Organization (NWO)

Publication of this thesis was financially supported by Wageningen University, Wageningen, The Netherlands.

Front cover: Some of the statistical parameters used during development of QSAR models in this thesis

Back cover: The chemical groups covered by the QSAR models developed in this thesis

Cover design & thesis layout: Ednah and Elton Zvinavashe

Printed by Ponsen & Looijen, Wageningen, The Netherlands