research for
man and environment

RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEU
NATIONAL INSTITUTE OF PUBLIC HEALTH AND THE ENVIRONMENT

RIVM report 601516 003

Implementation of QSARs in ecotoxicological risk assessments

R. Posthumus and W. Slooff

January 2001

This investigation has been performed by order and for the account of Directie RIVM, within the framework of project 601516/01/QS, ”Risk assessment of chemicals: Science and market”.

RIVM, P.O. Box 1, 3720 BA Bilthoven, telephone: 31 - 30 - 274 91 11; telefax: 31 - 30 - 274 2971

Abstract

Quantitative Structure Activity Relationship (QSAR) modelling techniques are overviewed here, along with descriptors which can be used in QSAR equations and the different statistical methods suitable for deriving QSARs. Discussed is the current state of the art on the use of QSAR estimates within the different regulatory assessment frameworks of the Centre for Substances and Risk Assessment at the Dutch National Institute of Public Health and the Environment (RIVM/CSR). Special emphasis has been placed on environmental effect assesssments. Commonly accepted QSAR equations and software programs used in environmental exposure assessments are described.Generally accepted QSAR equations and other QSARs for environmental effects documented in the literature are presented in tabular form; compounds acting by non-polar narcosis, polar narcosis, reactive chemicals and chemicals with a specific mode of action are presented separately. Recently developed computerised QSAR programs are also overviewed. Several recommendations are made for the use of QSARs within RIVM/CSR risk/hazard assessment frameworks, with experiences related to routine application of QSARs in effect assessments to be evaluated and reported on within one year.

Summary

The aim of this report is to instigate guidance in using QSARs (Quantitative structure-activity relationships) in the ecotoxicological risk assessment by RIVM/CSR. In a first RIVM-guidance document in 1992 their use has been advocated in case environmental data were lacking. However, recent inventories of bottle-necks in risk assessments of chemical substances indicated that in spite of their potential QSARs were hardly used in the assessments framework.

In the present report the QSAR approach is considered from a viewpoint of (scientific) supply and (policy) demand.
For environmental exposure assessment QSAR estimating procedures are widely accepted and QSAR estimates are commonly used for estimating physico-chemical parameters in exposure assessments. Therefore this report is mainly directed towards the derivation and application of QSARs in effect assessment. In this respect the following observations have been made:
Supply: Classification of chemicals according to their mode of action and the derivation of reliable QSARs is very complex. These days there is an array of potentially useful descriptors, models and statistical techniques of which the new approaches seem to be most promising. However, most of them still are in a developmental stage, and when available, they will require a high level of expertise for implementation. Only a limited number of QSARs for a few processes and/or chemical classes is directly available for use.
Demand: The policy needs for using QSARs in environmental effect assessment by RIVM/CSR seem to be limited. In most frameworks experimental toxicity data will be available and preferred. Only for existing substances more effort may be required in the QSAR approach, especially with respect to INS- and intervention values for sanitation and PTB-criteria for hazard assessment. However, in the latter case it is considered the responsibility of the industry to perform PTB-exercises: RIVM/CSR may be only involved in the assessment methodology (which may include selecting QSARs).

Based on both scientific and business/economical criteria it is recommended:

- To exclude QSAR development from CSR-activities, unless specifically commissioned by governmental authorities (based on supply/demand ratio).
- To apply QSARs in ecotoxicological (effects) assessment whenever reliable QSAR estimates are available. QSARs are not considered as circumstantial evidence only, but their use will contribute to a better understanding of toxicological modes of action and therefore will ultimately lead to better effect and risk assessments.
- To this end, to use for environmental effect assessment the computer program ECOSAR, which was found the most easy to handle computer program available.
- To improve the quality of the QSAR estimates from ECOSAR by excluding those QSAR estimates that have been based on less than 5 experimental data or with a r^2 beneath 0.7.
- If the chemical in question has more than one functional group, ECOSAR gives QSAR estimates for each of these groups. It is recommended to use the QSAR that generates the lowest value.

- To realise that the most reliable QSARs are available for chemicals with a non-specific mode of action (baseline toxicity); also for chemicals which acting by polar-narcosis a number of reliable QSARs has been developed. All CSR co-workers are considered to have basic knowledge about the application of these QSARs. In contrast, the number of QSARs available for reactive and specific acting chemicals is very limited, and their use for RIVM/CSR purposes is not considered to be of great interest .
- If no QSARs are available in ECOSAR, expert judgement is required for applying other published QSARs as presented in this report.
- To follow the newest developments in QSAR modelling techniques; models with high predictive power may become available and may be suitable for use within RIVM/CSR. However, application of these models probably requires specific knowledge, and special training of some CSR co-workers is expected to be necessary to use these programs in future.
- To apply routinely QSARs in (effects) assessments of all legislative frameworks and to evaluate and report after several months the experiences in the strategic research RIVM program.

Samenvatting

Dit rapport is bedoeld als leidraad voor het gebruik van QSAR's (kwantitatieve structuur-activiteitsrelaties) in de ecotoxicologische risicobeoordeling als uitgevoerd binnen RIVM/CSR. In een eerste RIVM leidraad uit 1992 werd het gebruik van QSAR's gepropageerd in geval milieugegevens ontbraken. Echter, een recentelijke inventarisatie van knelpunten bij het maken van risicobeoordelingen toonde aan dat, ondanks hun veelbelovende mogelijkheden, QSAR's nauwelijks gebruikt worden in de verschillende beoordelingskaders.

In het rapport wordt de QSAR-benadering beschouwd vanuit het oogpunt van (wetenschappelijk) aanbod en (beleidsmatige) vraag.
Bij de beoordeling van de blootstelling via het milieu behoren QSAR-schattingen tot de algemeen geaccepteerde procedures en het toepassen van QSAR's voor de schatting van physisch-chemische parameters is gebruikelijk. Daarom is dit rapport meer toegespitst op de afleiding en toepassing van QSAR's voor de beoordeling van effecten in het milieu. Met betrekking hierop kon het volgende worden waargenomen:
Aanbod: Klassificatie van chemische stoffen op grond van hun werkingsmechanisme en afleiding van betrouwbare QSAR's is een complex geheel. Er is tegenwoordig een scala aan mogelijk heel bruikbare descriptoren, modellen en statistische technieken, waarvan de allernieuwste benaderingen erg veelbelovend lijken. Een groot deel verkeert echter nog in ontwikkelingsfase. Indien ze beschikbaar komen is waarschijnlijk een hoge graad van expertise nodig om ze daadwerkelijk te kunnen toepassen. Slechts een beperkt aantal QSAR's voor enkele processen en/of chemische klassen kan direct worden toegepast.
Vraag: De beleidsmatige behoefte QSAR's te gebruiken voor de schatting van effecten in het milieu lijkt binnen RIVM/CSR erg beperkt. Voor de meeste beoordelingskaders zijn experimentele gegevens voorhanden en verdienen deze experimentele gegevens de voorkeur. Alleen voor bestaande stoffen zou de QSAR benadering daadwerkelijk kunnen worden toegepast, vooral met betrekking tot interventiewaarden en PTB-criteria. In het laatste geval echter ligt de verantwoordelijkheid voor de bepaling van de PTB-profielen bij de industrie; RIVM/CSR is alleen betrokken bij de ontwikkeling van de methodologie voor de risicobeoordeling (hetgeen wel de selectie van QSAR's zou kunnen inhouden).

Gebaseerd op zowel wetenschappelijke als ook zakelijke/economische criteria wordt het volgende aanbevolen:

- Ontwikkeing van QSAR's niet tot de CSR-activiteiten te rekenen, tenzij dit speciaal overheidswege wordt opgedragen.
- QSAR's toe te passen in de ecotoxicologische (effect)beoordeling indien betrouwbare QSAR's beschikbaar zijn. QSAR's niet te zien als alleen maar een bijkomstige ondersteuning van experimentele gegevens, maar ook als een bijdrage aan het verkrijgen van een beter inzicht te verkrijgen in het werkingsmechanisme van chemische stoffen (wat uiteindelijk zal kunnen leiden tot een betere risicobeoordeling).
- Met het oog op voorgaand punt, het computer programma ECOSAR te gebruiken dat het meest gebruikersvriendelijke programma is gebleken.

- De kwaliteit van de QSAR-schattingen met behulp van ECOSAR te vergroten door de QSAR's met een r^2 kleiner dan 0.7 en/of gebaseerd op minder dan vijf experimentele gegevens niet te gebruiken.
- In die gevallen dat een chemische verbinding meer dan één functionele groep bezit en ECOSAR meerdere QSAR schattingen geeft, de laagste QSAR-waarde te kiezen.
- Zich te realiseren dat de meest betrouwbare QSAR's beschikbaar zijn voor stoffen die een niet-polair of polair werkingsmechanisme vertonen; alle CSR-medewerkers worden geacht deze QSAR's te kunnen toepassen. Het beperkte aantal QSAR's dat aanwezig is voor reactieve en specifiek werkende verbindingen wordt niet van belang geacht voor toepassing binnen RIVM/CSR.
- Indien geen betrouwbare QSAR's met behulp van ECOSAR kunnen worden gevonden, is specifieke kennis en ervaring vereist om de andere in dit rapport gepresenteerde QSAR's toe te kunnen passen.
- De nieuwste ontwikkelingen op het gebied van QSAR modelleringstechnieken te blijven volgen. Misschien komen modellen met een hoge voorspellingsgraad beschikbaar die geschikt zijn om binnen RIVM/CSR toe te passen en kunnen enkele CSR-medewerkers opgeleid worden om deze modellen te gebruiken.
- Binnen alle verschillende risicobeoordelings kaders QSAR schattingen voor effect beoordelingen standaard toe te passen en na verloop van tijd de bevindingen ten aanzien van het gebruik van QSAR's te evalueren en te rapporteren in het strategisch onderzoeks programma van het RIVM.

Contents

1. Introduction

1.1 Background

In 1992 a first RIVM-guidance document on ecotoxicological risk assessment procedures for environmental chemical pollutants was published (Slooff, 1992). In this report the use of Quantitative Structure-Activity Relationships (QSARs) was proposed in case environmental data were lacking. For exposure assessment QSARs were proposed to calculate water/air partition coefficients, vapour pressure, adsorption coefficients, solubility, melting point, etc. In the effect assessment the use of QSARs (instead of results of laboratory toxicity tests!) was advocated for those chemicals that are classified as inert organic chemicals without specific mode of action (acting by non-polar narcosis) according to Verhaar et al. (1992) en Van Leeuwen et al. (1992). In the preliminary effect assessment QSAR information was recommended in case a group of structure related compounds is to be considered and QSAR estimates are available. Further physico-chemical characteristics (high Kow, low molecular weight) were considered as indicators for bioaccumulative potential.
Five years later an inventory of bottle-necks in the methodology of risk assessment of chemical substances indicated, among others, that in spite of their potential QSARs were hardly used in the assessments framework (Luttik and De Bruijn, 1998). As an example of their potential use the application in estimating (an)aerobic biodegradation of substances was mentioned. In Luttik and De Bruijn (1998) it was recommended to consider this gap between scientific knowledge and policy implementation with high priority. Interviews were held among experts and policy makers at the national level (Sijm and Luttik, 2000): once again a broader application of QSARs in ecotoxicological risk assessment was plead for with high priority. Also in an international brainstorm session on the development of a future “chemicals” strategy for the European Union (EU, 1999) research on QSARs was advocated to speed up assessments.

1.2 Goal

The goal of this report is to describe available QSARs and types of QSARs for ecotoxicological endpoints and to identify (further) possibilities on the use of QSARs in the (regulatory) ecotoxicological assessments of chemicals. The following questions are addressed:

- which QSARs are currently used at RIVM/CSR in the various ecotoxicological risk assessment frameworks and how are they applied?
- taking into account recent developments, which QSARs are to be considered for ecotoxicological risk assessment at CSR?
- which criteria should be applied to select QSARs for use within the risk assessment process by RIVM/CSR?

The report aims to instigate guidance in using QSARs in the ecotoxicological risk assessment by RIVM/CSR.

1.3 Method

A literature search was performed in the bibliographic database TOXLINE PLUS on CD-ROM (1995 to 2000). Relevant articles were collected and a retrospective search was performed by screening reference lists. All ecotoxicological QSAR data were evaluated, especially those referring to effect assessment and concerning chemical classes for which no or few QSARs were known. Expert sessions were held to discuss results and application possibilities.

1.4 Structure

In chapter 2 some basic background information is presented on QSARs, their derivation methods and characteristics. Chapter 3 focuses on the current use of QSARs in ecotoxicological assessment of various classes of chemicals (existing chemicals, new chemicals, plant protection products and biocides, veterinary medicinal products) as well as in the framework of setting standards (environmental quality standards, intervention values) and classification/labelling purposes. Chapter 4 presents the current availability of QSARs for environmental exposure and effect assessment and the availability of computerised QSAR models. Finally, chapter 5 evaluates the foregoing information and presents a proposal for QSAR application within RIVM-CSR.

2 Quantitative structure-activity relationships (QSARs)

2.1 Definition of QSARs and their derivation and use

QSARs are used to recognise and utilise the systemic relationships between the principal properties of chemicals and their biological and ecotoxicological activity. The principle of QSARs consists of relating the activities observed for a series of chemicals to a set of theoretical parameters, which are assumed to describe the relevant properties of their structures quantitatively. Derivation and application of QSARs requires three essential prerequisites:

- a) descriptors of the chemical *structure*
- b) measures of the chemical *activity*
- c) statistical techniques to quantify the *relationship*

For recent overviews on prerequisites for QSAR derivation and application reference is made to Nenzda, 1998; EU-DG-XII Project, 1995; ECETOC Technical Report, 1998.

2.2 Descriptors

Descriptors of the chemical structures are used to characterise and quantify properties of the entire molecule or substructural fragments or the physico-chemical properties of the chemical.

Three basic processes are involved in producing effects by chemicals in the environment:

- Transport and partitioning, which is predominantly governed by the distribution of the chemical between the hydrophobic and the aqueous phase.
- Formation of the chemical/target complex, which depends on the forces of interaction between the two components, mostly hydrophobic bonding, their charge, their shape and their size.
- Reactions of the chemical/target complex, which are often controlled by the compounds electron distribution.

There are two kinds of descriptors that describe the properties of a chemical related to these processes:

- *Quantitative* descriptors, which are based on molecular structures as represented by constitutional formulas. An unambiguous definition of structures is required, also with regard to geometry, conformation, isomerism, tautomerism, stereochemistry, etc.
- *Structura*l descriptors which have to represent the entire spectrum of the molecular or submolecular interactions leading to a particular property (i.e. hydrophobic, polar, electrostatic, and steric interactions).

Thus, basically the effects of chemicals on environmental systems may be caused by three principal properties: the lipophilic, electronic and steric features of the compound.

The following descriptors can be distinguished:

2.2.1 Chemical descriptors

Hydrophobicity has emerged as the key parameter for assessing the potential environmental impact of contaminants. This property is a measure for the preference of a substance for partitioning into either aqueous or non-aqueous phases. The partitioning between compartments of different polarity determines the rate and direction of the transport of chemicals in the environment and thus their accumulation in some of its components:

Partition coefficient (P) = equilibrium concentration in non-aqueous phase/concentration in water

The 1-octanol/water system is accepted as the general reference system.

Substances of high hydrophobicity are usually small molecules with polar, possibly charged, moieties. If P approaches zero, the chemical will be so insoluble in non-polar (fatty) phases that it is unlikely to permeate lipid membranes and it will remain localised in the first aqueous phase it contacts. Conversely, as P approaches infinity, the chemical will be so insoluble in water that it tends to be trapped in non-polar phases, where it accumulates.

The partitioning between aqueous and non-aqueous phases in the environment concerns biotic and abiotic phases.

2.2.2 Electronic descriptors

Electronic parameters are used to account for ionisation and dissociation of chemicals. Parameters that describe the electron density distribution of molecules should only be used in QSAR studies if specific, well-defined, interactions are modelled and the respective features of both reaction partners are known. The mutual interactions due to electrostatic interactions (ion-ion interactions, ion-dipole interactions, dipole-dipole interactions) and hydrogen bonding may control the extent and strength of intermolecular chemical/target interactions. The following electronic descriptors have been described:

- The negative decadic logarithm of the acid/base dissociation constant *pKa* is an electronic descriptor which can be used to correct the logKow for the fractions of ionised/unionised species.
- *Hammett substituents constants* σ is used in an equation to formalise the electron-withdrawing/releasing potency of substituents on a aromatic system. The dimensionless σ values encode the substituent constants that express the polar effect of the substituents relative to hydrogen. The σ values are generally assumed constant for a given substituent, but they vary with the substitution pattern on the parent compound; σ values can be retrieved from tabulations. If several substituents are present on a parent structure, the sum of the corresponding σ values is used. The second application of σ values for QSARs in environmental sciences is for the estimation of the pKa of substituted aromatic compounds.
- The use of *hydrogen bonding*, intra- as well as intermolecular, is often restricted to an indicator parameter characterising either the presence or absence of hydrogen bond donor or acceptor functions. The respective solvatochromic (LSER, Linear Solvation Energy Relationship) parameters α (an estimation of acidity representing the ability of a solute to accept an electron and of a solvent to release an electron) and β (an estimation of basicity representing the ability of the solute

- molecule to release an electron and of a solvent to accept an electron) have to be determined empirically and hence are not universally applicable for making predictions.
- *Quantum chemical methods* may be applied to obtain a large variety of stereo-electronic descriptors such as ionisation potential, polarisability, electron affinity, dipole moment, charge densities, HOMO- (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital)-energies or atomic charges.
- Further ways of describing electronic features of chemicals comprise *spectroscopic data* (e.g. IR, UV or NMR measurements).

2.2.3 Geometric descriptors

Geometric descriptors are indicators for the shape and size of molecules and rule a variety of processes related to the fate and effects of environmental contaminants, such as size-limited diffusion, fit to target sites and shielding of reactive moieties. Membranes and other organic barriers are permeable only to compounds of an effective diameter <= 9.5 Å. However, it has to be noted that this rule holds only for passive uptake from water through the gills and not for uptake from the gastro-intestestinal tract. In addition, from experimental results it appeared that this effective diameter is species dependent. Larger molecules are hindered to diffuse through membranes and, unless active transport mechanisms act, they cannot reach intracellular regions. To be able to react with specific binding sites the chemical has to have a complementary constitution; the lesser the degree of fit the lesser the intrinsic activity. Bulky substituents near reactive centres in a molecule may shield this moiety and thus reduce its reactivity.

Size-related parameters may be represented by: molar volume V=(molecular weight/density); Van der Waals volume (V_{vdW}), TMV (Total Molar Volume), TSA (Total Surface Area), SASA (Solvent Accessible Surface Area), and SAVOL (Solvent Accessible Volume). Classical steric parameters are the Taft E_S substituent constants which reflect the shielding of a reactive centre by the substituent *R*. E_S constants are given in tabulations.

2.2.4 Topological descriptors

Topological descriptors reflect the connectivity between atoms in a molecule based on a graph theory. This approach is intended for planar structures.

Molecular connectivity parameters:

- connectivity indices, χ values, encode information from the molecular size at the lowest level of connection to implicit information about the three dimensional configuration at the higher levels of connection representing the complexity of the molecules.
- shape parameters, κ values, encode information about the shape of the molecule and encode a value relative to the maximum number of bonds in the isomeric star graph and the minimum number in the isomeric linear graph.

Connectivity indices can be calculated from a molecular formula (Landrum, 1999).

2.2.5 Polarisability descriptors

Polarisability-related parameters such as molar refractivity (MR) are mostly described as steric descriptors with the acknowledgement of some extra electronic information. Their dual character combines elements that account for the general steric bulk as well as for the polarisability, which refers to the displacement of charges within molecules influenced by an electric field (i.e. induced dipoles).

2.2.6 3D-descriptors

QSAR studies on molecular toxicology usually employ classical 2D descriptors to correlate the toxicity measures. However, in many molecular families the three-dimensional effects play an important role in their properties, and classical 2D QSARs become inappropriate to describe accurately these systems. Recently some three-dimensional approaches have been proposed to overcome these problems:

Molecular similarity techniques (MQSM) are based upon the premise that similar molecules possess similar properties. The concept “similar” can be defined in several ways, and it is not necessary to encompass descriptors related to the whole molecule, but only to a fragment or relevant topological aspect. Of all the molecular similarity approaches, Quantum similarity employs a fundamental quantum-mechanical descriptor, the first-order electronic density function, as a source for the resemblance quantification. Until now, MQSM application to derive QSARs has mainly been performed on chemicals of pharmacological interest. A first approach to the characterisation of molecular toxicity using MQSM was recently reported by Girones et al.(1999). The study was based, however, only on one descriptor, i.e. the electron-electron repulsion energy. Robert & Carbo-Doca (1999) approached molecular toxicity from the point of view of the entire similarity matrix, and depicts a possible alternative to classic QSAR approaches in toxicology.

A promising 3D QSAR descriptor is EVA (eigenvalue) in which calculated IR vibrational frequencies are used to derive high-dimensional descriptor vectors for predictive partial least-square models. EVA does not require the generation of structural alignment. Tuppurainen & Ruuskanen (2000) used a modification of EVA, EEVA (electronic eigenvalue) in which molecular orbital energies are used to derive QSAR descriptors instead of vibrational normal modes, for the prediction of Ah receptor binding affinities of PCBs, PCDDs, and PCDFs.The method seemed to be suitable for ‘pure’ electronic substituent effects, i.e. for cases in which both hydrophobic and steric factors are of minor importance.

Another approach is the WHIM theory (Weighted Holistic Invariant Molecular). WHIM descriptors contain information about the whole 3D-molecular structure in terms of size, shape, symmetry and atom distribution. These indices are calculated from (x, y, z)-co-ordinates of a 3D-structure of the molecule, usually from a spatial conformation of minimum energy, within different weighting scheme, in a straightforward manner and represent a very general approach to describe molecules in a unitary conceptual framework (Todeschini & Gramatica, 1997). A further description of 3D modelling techniques is given in chapter 4.3: Computerised QSAR models.

2.3 Measures of chemical activity

The activity measures in environmental sciences concern two kinds of parameters (endpoints):
- ecotoxicological effect parameters (e.g. LC50, NOEC)
- environmental fate- and exposure parameters (e.g logKow, biodegradation)

These two endpoints are of different nature and therefore, different techniques may be appropriate for the respective structure-activity analyses. QSARs for exposure-related, and effect-related parameters are described in Chapter 4. Models that describe physico-chemical parameters like logKow, logKoc. photolysis etc. are also called QSPRs (Quantitative Structure-Property Relationship); other abbreviations found in the literature are QSTR (Quantitative Structure-Toxicity Relationship), QSBR (Quantitative Structure-Biodegradability Relationship) and QSPR for Quantitative Structure-Permeability Relationship.

2.4 Modelling techniques and statistics

2.4.1 Modelling techniques

Three different QSAR modelling techniques can be distinguished:

- *Single descriptor models*: endpoint of a number of compounds is related to a single descriptor. Typically, the compounds are chemical related and have been assayed in a single run. Transferring such models to a distinct class of compounds at best yields a relationship with significantly different regression coefficients, or does not yield any satisfying regression at all. These models merely function as compact recordings of experimental data, rather than having predictive power for untested compounds.
- *Multiple descriptor models*: several relationships exist which are based on multiple linear regression (MLR). Descriptors: connectivity indices, Hammett constants, Taft parameters, van der Waals radii, electronic parameters, and partition coefficients. Most models contain less than five descriptors, which may be too modest for obtaining robust models.
- *Group contribution models*: In *linear contribution models* the biological or physico-chemical activity of compounds is predicted from the presence of structural fragments. A shortcoming of the linear group contribution method is that the fragment contributions are thought to be additive and the interactions between structural fragments are thus not explicitly accounted for. A further disadvantage of most group contribution models is related to the use of a discontinuous endpoint; e.g. for biodegradation models often a classification is made in readily or non-readily biodegradable. This precludes calculations of common statistics like for instance r^2. *The nonlinear group contribution models* are based on the application of neural networks (NN). In these models often structural fragments are used, and sometimes the models are extensions of linear fragment models. The idea behind NN-models is that the mutual interaction of fragments is taken into account. A distinct problem with neural networks is the lack of an analytical expression (Langenberg et al., 1996).

2.4.2 Statistical techniques

Various statistical techniques can be distinguished to quantify the relationships:

Regression analysis
Regression analysis is an exact mathematical procedure to correlate independent *X* variables (descriptors) with dependent *Y* variables (activity measures). Both physicochemical descriptors (Hansch approach) and substructure indicators (Free-Wilson method) may be used as independent variables:

- *Linear free-energy relationships: the Hansch approach*
  The Hansch approach uses multiple regression analysis to describe changes in the observed effects among chemicals from differences in their physico-chemical properties. The basic presumption is the linear dependence of the activity on the free energy (linear free-energy activity relationship: LFERs). The range of the descriptors covered by the compounds included in the model sets the limits for predictive application of the QSAR, because there is no certainty that the relationship holds beyond this domain. With logKow-dependent models, for example, it is frequently found that they are linear between logKow=0 and logKow=5, but reveal curvature outside this range.
- *Substructural models: the Free-Wilson method*
  This method uses multiple regression analysis to describe changes in the observed effects among chemicals from indicators of the presence or absence of certain substructures in their molecule. The activity of the parent compound is taken as the reference and the contributions of the substituents of the derivatives are accounted for by the indicator variables. Except for the molecular structure of the compounds and the respective activity measures, no further information is needed. Hence, the analysis can be conducted fast and simply. Its predictive use is strictly limited to substances with substructures in the data set used to derive the model.

It should be noted that a variety of factors may contribute to a non-linear dependence of activity:

- activity is measured before steady state has been reached
- solubility limits are exceeded
- micelles formation occurs for highly lipophilic compounds
- diffusion across membranes is hindered for large molecules
- transport to the target site involves partitioning between phases of different lipophilicity
- the model assumptions become invalid if a certain parameter value is surpassed.

Regression analysis is not applicable for modelling discrete response (e.g. readily biodegradable or non-readily biodegradable), because two-point correlations are the inevitable result.

Multivariate statistics
Multivariate statistics is required when large data arrays have to be handled on a multitude of endpoints and parameters. To recognise the extent and the nature of the basic properties encoded in such data sets, principal component analysis (PCA), factor analysis (FA) or partial least squares (PLS) can be used.

Principal components (PCs) are orthogonal vectors calculated by linear combination of the original variables, and they are supposed to represent substantial, but unobservable, properties. The first PC represents the maximum variance in the data cloud; the second PC is orthogonal to the first PC and represents the second largest variance in the data cloud; the third PC is orthogonal to the first and the second and represents the third largest variance in the data cloud. This procedure is continued until n orthogonal PCs are calculated from the n variables. All variables in a group have relatively small correlations with variables in other groups. Only those PCs are selected that represent a significant amount of the total information.

Partial least squares (PLS) analysis allows the simultaneous investigation of the relationship between a multitude of activity data (Y matrix) and a set of chemical descriptors (X matrix) through latent variables. The latent variables correspond to the component scores in the PCA and the respective coefficients to the PCA loading vectors. The PLS model can also be applied when the number of descriptors exceeds the number of compounds in the data set. Main difference between PCA and PLS: PCA is based on the maximum variance criterion, whereas PLS uses covariance with another set of variables (X matrix).

Classification methods

The analysis of semiquantitative activity data is not possible through classical regression techniques; in this case classification methods are required such as pattern recognition, cluster analysis, discriminant analysis and neural networks.

- *Pattern recognition* techniques can be used to identify the relationships between groups of compounds as well as to classify individual chemicals. The statistical methods used include PCA, FA, and PLS. Prior knowledge of the classes of some compounds (the training set) is needed as a basis for assigning further compounds to these classes. The efficacy of any classification model for predictions strongly depends on the appropriate selection of descriptors.
- *Cluster analysis* is an explanatory data analysis technique aimed at grouping items (e.g. chemicals and their properties) into clusters of similar items according to their position in the multidimensional parameter space.
- *Discriminant analysis* yields a function that separates the members of different classes based on either physico-chemical descriptors or indicator variables, such as the presence or absence of defined substructures. Successful discriminant analysis is based on the assumption that the data in each class are distributed normally and all classes have the same covariance matrix.
- *Neural networks*: A currently popular approach to classification and pattern-recognition problems involves neural networks (backpropagation method). Neural networks are mainly used as (non-)linear approximations to multivariate functions or as classifiers. Neural networks are in fact mathematical models of mutually interconnected functions arranged in one or usually three layers, which are supposed to produce a classifying output for a given multivariate input. Because of their black-box character, neural network models are hardly interpretable in mechanistic terms (Nenzda, 1998).

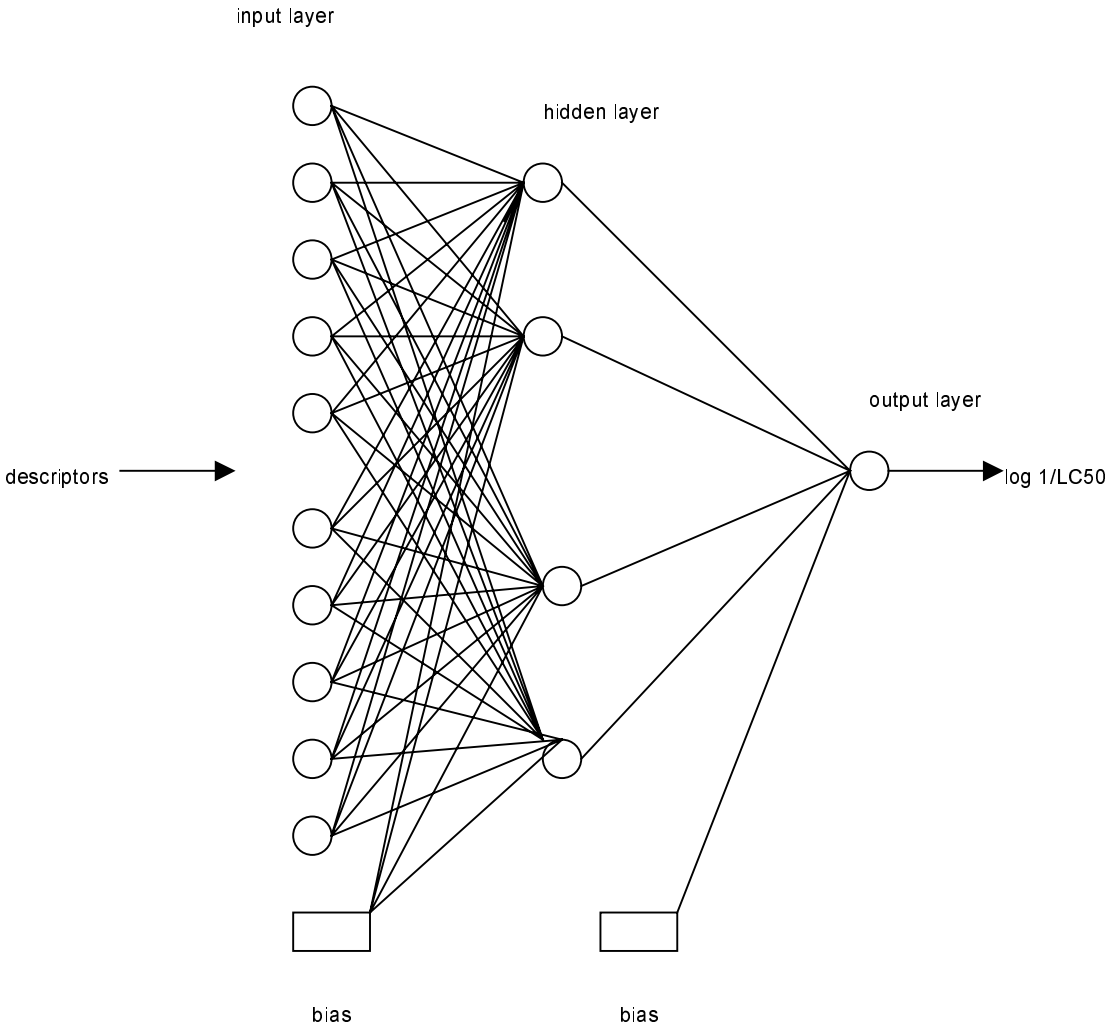


Figure 2.1: A three layer neural network (Devillers & Domine, 1999)

2.5 Conclusions

It can be concluded that two decades of QSAR-development have led to quite some descriptors that can be used to characterise the properties of a chemical, and has resulted in many modelling and statistical techniques to qualify and quantify the relationships between the descriptors and the environmental behaviour and impact of chemicals. Especially since the appearance of the first RIVM guidance document in 1992, there has been a massive flow of scientific publications on QSARs (figure 2.1). This increased scientific attention holds a promise for further implementation of QSARs in ecotoxicological risk assessment. Most promising are the more complex QSAR methodologies (3D-descriptors and neural network analysis) which demonstrated high predictive power and are not limited to a single chemical class.

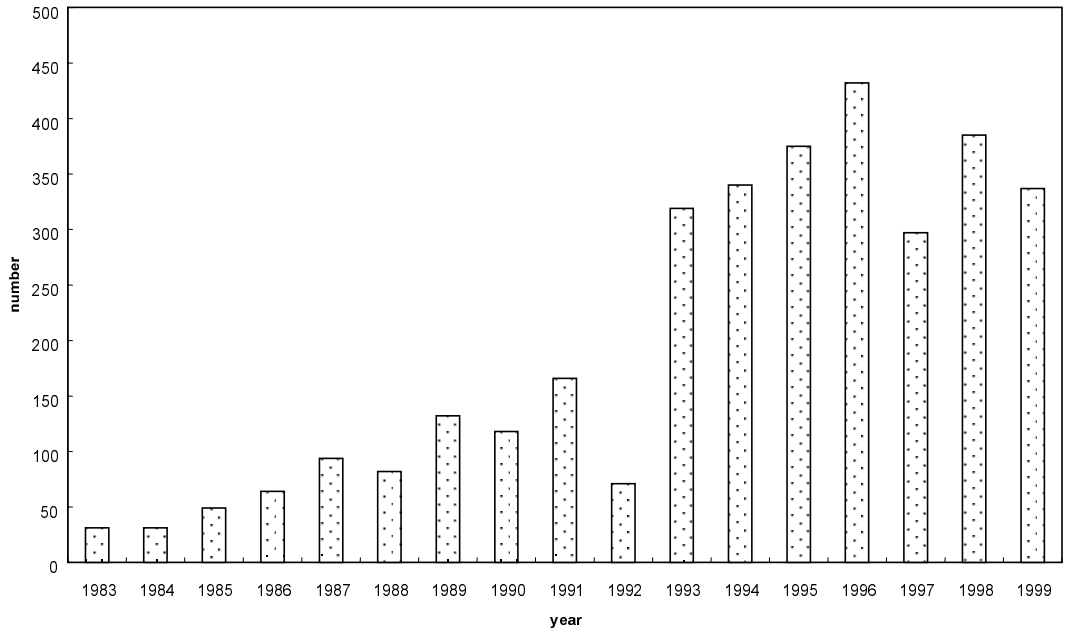


Figure 2.2: Number of publications on QSAR in the period 1983-2000 derived from online search in the database BIOSIS

3. Current use of QSARs in ecotoxicological assessments

Various assessment categories can be identified and for each of them legislation has been developed, as well as methodology to assess the hazard/risk of chemical substances, including the use of QSARs. These assessment categories may be based on the category of chemical (veterinary medicinal products, plant protection products and biocides, new chemicals, existing chemicals) or the type of policy values (integrated environmental quality standards, intervention values, classification).

3.1 Veterinary medicinal products

For registration, manufacturers of veterinary medicinal products have to provide the information required for risk assessment. In case the information is not sufficient or if the quality is not considered to be valid, the manufacturer is asked to supply the lacking or more valid data. In general, veterinary medicines are chemicals with a specific mode of action. For such chemicals very little or no QSARs are available up till now (see section 5). Therefore, QSARs are not used for this kind of environmental effect assessment. For environmental exposure assessment QSAR estimates may be used for physico-chemical parameters.
The rules governing medicinal products in the European Union are laid down in Eudralex (1999).

3.2 Plant protection products and biocides

For registration, manufacturers of plant protection products and biocides have to provide the information required for risk assessment. In case the information is not sufficient or the quality is not considered to be valid, the manufacturer is asked to supply the lacking or more valid data. Plant protection products and biocides are chemicals with a specific mode of toxic action or they belong to the group of reactive chemicals. Little or no QSARs are available for such chemicals up till now (see section 5). Therefore, QSARs are not used for this kind of environmental effect assessment. For environmental exposure assessment QSAR estimates may be used for physico-chemical parameters and BCFs.
The Council of Europe and the EPPO (European Mediterranean Plant Protection Organisation) published decision-making schemes which can be used in effect assessments. At the national level the primary responsibility for risk assessment for plant protection products and biocides is laid down at the Dutch Board for Authorisation of Pesticides (“CTB”). The procedure and methodology of risk assessment, originally developed by RIVM, was agreed upon by an Interdepartmental Steering Group (INS) and the CTB and accepted by the Steering Group Pesticide Policy.

3.3 New chemicals

Under the Dangerous Substances Directive, new chemicals have to be notified before marketing: the basic data set (identification, physicochemical, toxic and ecotoxic properties) is provided by the notifier. If there are questions about the provided information the notifier is asked for explanation or further data. Experimentally derived data are sufficiently available and QSARs are not used for risk assessment. The risk evaluation is performed according to the Technical Guidance document (EU, 1996).

3.4 Existing chemicals

According to the EU regulation on the risk assessment of existing chemicals: producers and importers are obliged to complete HEDSETs with all available data. QSARs are used for estimating lacking physicochemical parameters in environmental exposure assessment. Rules for the use of QSARs in environmental effects assessment are given in the Technical Guidance Document (EU, 1996): see 3.6.1 and chapter 4. Although the use of QSARs is propagated, in practice this possibility is seldom utilised in environmental effect assessment by CSR.
Currently two opposite developments can be distinguished in the approach towards regulation of substances, one focuses on the high priority pollutants, the other on controlling the risks of the non-assessed chemicals:
One development is related to the defensibility and justification of environmental adverse effect levels in the process of balancing economic/ environmental costs and benefits of (high) priority compounds. In this context there is a movement to true risk assessments – often region or site specific - with a measure of the probability and severity of an adverse effect, characterised in terms of statistical distribution with a most probable value for the risk accompanied with a confidence interval (SSD-boek, 2000). Since QSAR based predictions include inherent uncertainty, their use in assessing actual risk and damage will be very limited.
The other development is related to the problem of the lack of data on the majority of the existing chemicals. At present the high production volume chemical (HPVCs) are being assessed. For these (ca. 2700) substances relatively many experimental data are available. Yet there are significant gaps in what is known in the open literature: information on biodegradation and Kow is only available for around a third of these substances, bioconcentration factors for 15% and terrestrial toxicity for 3% (EU, 1999). These days a start has to be made with assessing the low production volume chemicals (LPVC), the estimated number ranging up to 30,000. For this group of substances the data availability is much smaller. In this scope the Dutch 3rd National Environmental Policy Plan (NMP, 1998) announced that new approaches towards regulating substances will be developed, in co-operation with European partners. In this framework a project has been initiated (Strategy On Management of Substances: SOMS) in which fast screening methods for hazard identification play an important role in tackling the thousands of chemicals of which data are lacking. For this purpose PTB-criteria (persistence, toxicity, bioaccumulative potential) are considered useful, besides other criteria like MCR (mutagenicity, carcinogenicity, reproductiontoxicity). Earlier, the ministry of VROM used the PTB-approach in priority setting of existing

chemical substances which was originally proposed in the seventies (Canton and Slooff, 1979). From recent studies on methods for determining PTB-properties it is clear that QSARs are considered as an important tool (BKH/Haskoning, 1998; Sijm et al., 1999; Hansen et al., 1999). Hence, there is a future for QSAR applications in this development. However, according to the view on future chemical policy the industry should provide the authorities with PTB-information and it is uncertain which role RIVM will play in screening these data.

3.5 Setting of Integrated Environmental Quality Standards and Intervention values

In this framework the standards are based on data from public literature. Often these data are scarce and QSARS may be useful to fill in the data gaps. The procedure to derive Maximum Permissible Concentrations (MPCs) and Ecotoxicological serious Soil Contamination Concentration (ECOTOX SCCs) has been described by Slooff (1992), Crommentuijn et al (1997) and Kalf et al. (1999): see chapter 3.6.3.

3.6 Guidance documents and systems

3.6.1 Technical Guidance Document (TGD)

International guidance in risk assessment methodology for new and existing chemicals is given in the EU-framework in the Technical Guidance Document (TGD, 1996).
In the TGD (1996) all relevant QSARs are considered, but they are considered as *supporting tool* in taking the decision whether test results are valid for use in risk assessment.
For establishing an exposure level or exposure concentration based on modelling, several physico-chemical/fate parameters are used. In the absence of experimental data, e.g. because it is not possible to obtain reliable measured data, these parameters *must* be derived by applying QSARs. With respect to effect assessment, however, the TGD is rather reticent about QSAR application.
The QSARs proposed in the TGD for environmental exposure and effect assessment are described in chapter 4. At present the Guidance Document is in the process of updating, but no efforts will be taken to update the progress in QSAR knowledge (De Bruijn, pers. com.).

3.6.2 Uniform System for the Evaluation of Substances (USES)

The procedure and methodology of risk assessment of chemical substances (new and existing substances, agricultural and non-agricultural pesticides) as performed by RIVM/CSR is set down in the risk assessment tool USES 3.0 (1999). As required, this tool is attuned to current chemical management policies and in accordance with the principles laid down in the TGD (1996) for new and existing chemicals and the Uniform Principles for Pesticides of the European Commission. In USES 3.0 the following statements are made:

- For the chemical substances of concern data sets will be available (minimally bases-set). Secondary chemical-specific data such as partition coefficients and bioconcentration factors will be scarce. These data gaps will be filled by estimated data, using generally agreed procedures like QSARs, or by default values. In principle, QSAR estimation of parameter values is preferable to use above default values in those cases that default values lead to a non-realistic worst case.
- QSARs have been selected that are simple, are applicable to a wide range of substances and that estimate secondary data on the basis of available physico-chemical descriptors of compounds. Of the various QSARs discussed in the TGD (1996), a QSAR for Koc and BCF is included in USES. No other QSARs have been implemented in (E)USES, but any results can be entered manually (yet a factor of 10 on the QSAR derived long-term NOEC is used in the Hydrocarbon Block Method (HBM) for petroleum products).

3.6.3 The use of QSARs within RIVM/CSR

The procedure to derive Maximum Permissible Concentrations (MPCs) and Ecotoxicological serious Soil Contamination Concentration (ECOTOX SCCs) has been described by Slooff (1992), Crommentuijn et al (1997) and Kalf et al. (1999). At first glance, these reports indicate that the view on the use of QSARs by RIVM has changed in the last decade:
According to the RIVM Guidance document of Slooff (1992) QSARs may be used to derive maximum permissible concentrations (MPCs) for chemicals with the same mode of toxic action (e.g. chlorobenzenes, chlorophenols) if at least for one chemical within such a group a MPC based on experimentally determined NOEC-values is available. The MPC for the other chemicals in the group can be subsequently derived by applying the ratios between the QSAR-values to the HC5 of that chemical. According to the guideline for deriving environmental quality objectives (Crommentuijn et al., 1997; SOP CSR/H/003) QSARs may be used when the modified EPA method (OECD, 1992) is applied for derivation of the MPC. However, in the latest protocol for derivation of harmonised MPCs (Kalf et al, 1999) the use of QSARs for generating toxicity data is no longer propagated. Upon inquiry it appeared that it was not intended to leave out QSAR application for effects assessments, the protocol being focused on incorporating the use of assessment factors of the ECB (1996) and views of the CTB (van Wezel, pers. com.).

3.7 Conclusions

For environmental exposure assessment QSAR estimates are commonly used for estimating physico-chemical parameters. In contrast, the use of QSARs in environmental effect assessment appears to be minimally in spite of all investigations made in the last decade to develop reliable QSARs. The most important reason is that in most legislative frameworks the manufacturer has to supply (additional) experimental data, if required, and experimentally derived data are generally considered more reliable than QSAR estimates and therefore are preferred. Another reason is that no QSARs are available for chemicals like pesticides. Application of

QSAR for effect assessment therefore may be restricted to derive integrated environmental quality standards and intervention values of existing chemicals only. Although estimating PTB-profiles for classification purposes is often QSAR-based, in the view of the new development in the chemical policy the industry has to provide these profiles and so far it is unknown whether RIVM/CSR will be consulted or involved in this process. Hence, from the perspective of legislation and chemical substances policy there is no urgent need for implementing QSAR application in RIVM/CSR working processes.

4. Availability of QSARs

In this chapter the QSARs proposed in the TGD for environmental exposure assessment are described, together with QSARs which have been reviewed and found to be reliable in the EU Project (1995), Nenzda (1998) and ECETOC (1998). QSARs proposed in the TGD and the EU Project for environmental effect assessment are listed in Table 1 to 4. Other QSARs found for environmental effect assessment are added to the tables; it should be noted, however, that none of these added QSARs have a generally approved status.

4.1 QSARS for environmental exposure assessment

4.1.1 Octanol/water partition coefficient logKow

The log Kow is the key parameter for assessing the potential environmental impact of contaminants. The partitioning between compartments of different polarity determines the rate and direction of the transport of chemicals in the environment and thus their accumulation in some of its components.
The logKow for ionisable substances depends on pH of the environment and on the pKa of the substance. To correct for ionisability the following equations are given:

$\log Kow\ (corrected) = \log Kow - \log(1 + 10^{pH-pKa})$ for acids
$\log Kow\ (corrected) = \log Kow - \log(1 + 10^{pKa-pH})$ for bases

Estimation of logKow values is based on the assumption that the hydrophobicity contributions of the different fragments in a molecule are additive (Hansch and Leo, 1979; Rekker, 1977, Lyman, 1990)
Software: CLOGP (Medchem,1989), ASTER (US EPA), AUTOLOGP (Devillers et al., 1995) and KOWWIN (SRC,1997). KOWWIN uses a fragment constant methodology; coefficients for 135 individual fragments and groups, and 255 correction factors were derived by multiple regression of more than 2400 reliable measured logKow values and comparisons between predicted and experimental values (Sijm et al.,1999)
Domain: Group contribution methods are limited to the calculation of logKow values for inert organic compounds, and they cannot be reliably applied for inorganic compounds, surface-active compounds, chelating compounds, organometallic compounds, partly/fully dissociated compounds, compounds of extremely high or low lipophilicity, and mixtures (e.g.impure compounds). LogKow values between –3 and 7.

4.1.2 Octanol/air partition coefficient (Koa)

The n-octanol/air partition coefficient (Koa) is a key parameter in studies on environmental fate of (non-ionic) organic substances. LogKoa is used in the estimation of the extent of volatilisation from or sorption to soil and vegetation. Although there are some experimental methods available, usually Koa is estimated from the ratio of Kow and HE: Koa = Kow/HE (HE: see 4.1.5). (Sijm et al., 1999).

4.1.3 Water solubility (WS)

The aqueous solubility is a key parameter in studies on environmental fate (and toxicity) of chemical substances. Most of the highly water-soluble substances have low logKow values, low adsorption on soil and sediment, low bioconcentration and they tend to be readily biodegradable. QSARs models for water solubility with logKow as sole descriptor are numerous (Chiou et al. (1977), Hansch et al. (1968), Isnard & Lambert (1989); Valvani et al. (1981), Müller & Klein (1992), Kenaga and Goring, 1980) These logKow dependent QSARs are principally applicable for liquids not for solids. For solids a correction has to be made in the model for lattice energy via melting point (Nenzda, 1998). See table 4.1. for QSAR models estimating water solubility.

Table 4.1: QSAR models for estimating water solubility (S in mol/l): linear log S / logKow correlations, (1-16) including melting point correction (T_m in ºC, 17-21) and correlations with various parameters (22-24) (Nenzda, 1998)

	Chemical class	Equation	R2	n	logKow range
1	Aromatic P-containing Pesticides	Log S = - 1.49 logKow + 1.46	0.98	34	1-7
2	Phosphoric acid esters	Log S = - 0.93 logKow + 2.2	0.95	11	0-5
3	Alcohols	Log S = - 1.11 logKow + 0.93	0.97	41	0.6-2.8
4	Ketones	Log S = - 1.23 logKow + 0.72	0.98	13	0.3-2.8
5	Esters	Log S = - 1.01 logKow + 0572	0.99	18	0.2-4.7
6	Ethers	Log S = - 1.183 logKow + 0.94	0.94	12	0.8-2
7	Alkyl halides	Log S = - 1.22 logKow + 0.83	0.93	20	1.4-3
8	Alkynes	Log S = - 1.29 logKow + 1.04	0.95	7	2-4
9	Alkenes	Log S = - 1.29 logKow + 0.25	0.99	12	1.7-3.7
10	Aromatics	Log S = - 1.00 logKow + 0.34	0.98	16	0.9-3.7
11	Alkanes	Log S = - 1.24 logKow – 0.25	0.95	17	2.3-4
12	Diverse liquids	Log S = - 1.34 logKow + 0.98	0.93	156	0-5
13	Liquids, solids	Log S = - 1.38 logKow + 1.17	0.97	300	0-8
14	Diverse liquids	Log S = - 1.02 logKow + 0.52	0.92	111	-1-5
15	Diverse liquids	Log S = - 1.16 logKow + 0.79	0.97	156	0-5
16	Pesticides, halogenated benzenes, PAHs, P-containing pesticides	Log S = - 0.92 logKow + 1.18	0.86	90	-4-6.5
17	PAHs	Log S = - 0.88 logKow – 0.01 T_m – 0.01	0.99	32	3-7
18	Halogenated benzenes	Log S = - 0.99 logKow – 0.01 T_m + 0.72	0.99	35	2-6.5
19	Halogenated aromatics	Log S = - 1.37 logKow – 0.0094 T_m + 1.64	-	15	-
20	Liquids, solids	Log S = - 0.94 logKow – 0.01 T_m + 0.32	0.98	162	2-8
21	Liquids, solids	Log S = - 1.26 logKow – (0.0054 T_m – 25) + 1.0	-	300	0-8
22	Not for stronger hydrogen-Bonding donor solutes	Log S = - 3.36 V + 0.46 π* + 5.23 β + 0.55	0.99	93	-
23	Diverse	Log S = 1.63 $^0\chi$ – 1.37 $^0\chi^v$ + 1.00 Φ + 1.56	0.99	470	-
24	Diverse	Log S = x^w / (1 - x^w)	-	-	-

T_m: melting point
π *: solute/solvent dipolarity/polarizability
β : hydrogen bond acceptor basicity
$^0\chi$, $^0\chi^v$: zero-order connectivity indices
Φ : polarizability
x^w : mole fraction

Software: The WSKOW (SRC, 1997) program estimates the aqueous solubility of organic compounds with logKow, molecular weight, and melting point (if available) as input parameters, and corrections for 15 structure types.
Domain: (mainly non-ionic) organic substances. Not applicable for metals, metalloids and organometals.

4.1.4 Vapour pressure (VP)

Partitioning and transport of substances between environmental compartments are determined by vapour pressure (VP). Highly volatile chemicals show rapid, long-distance dispersion in the atmosphere, and may reach water, soil and terrestrial animals and plants by dry and wet deposition. Several methods are available for calculating VP based on derivations of the Clausius-Clapeyron equation (equation 1 and 2: Grain, 1990; equation 3: Lyman, 1985; equation 4: Altschuh et al., 1993; equation 5: Mackay et al., 1982; equation 6: Mishra & Yalkowski, 1991).

Table 4.2: QSAR models for estimating vapour pressure (Nenzda, 1998)

	chemical class	equation
1	liquids	$\ln VP\ (atm) = K_F\ (8.75 + R \ln T_b)\ (T_b - C)^2\ /\ (0.97\ RT)\ [1\ /\ T_b - C) - 1\ /\ (T - C)$
2	liquids, (solids)	$\ln VP\ (atm) = K_F\ (8.75 + R \ln T_b)\ /\ (0.97\ RT)\ [1 - (3 - 27^*)^{m^*} / T^* - 2m(3 - 27^*)^{m-1}\ \ln 7^*]$
3	liquids, solids	$\ln VP\ (atm) = K_F \ln (RT_b) / 0.97)\ [1 - (3 - 27^*)^m / T^* - 2m^*\ (3 - 27^*)^{m^* - 1}\ \ln 7^*] + \Delta p_{(s)}$ $\ln \Delta p_{(s)} = 0.6 \ln (RT_m)\ [1 - (3 - 27^{**})^{m^*} / T^{**} - 2m^{*}\ (3 - 27^*)^{m^* - 1}\ \ln 7^{**}]$
4	liquids, solids	$\log VP\ (Pa) = 0.434\ (\Delta S_v / R)\ T_b\ \Phi(k, T_b, T) + f(T_m) + 5.006$ $\Delta S_v = 1.03\ [10.60 + 3.66 \log T_b + (1 / MW)(0.0935\ T_b + 1.035 \times 10^{-3}\ T_b^2 - 1.345 \times 10^{-6}\ T_b^3)]$ $\Phi\ (k, T_b, T) = (1 + K)\ (1 / T_b - 1 / T) + (K / T_b) \ln (T_b / T)$
5	liquids,solid hydrocarbons, halogenated hydrocarbons	$\ln VP\ (atm) + - (4.4 + \ln T_b)\ (1.803\ (T_b / T - 1) - 8.03 \ln (T_b / T) - 6.8\ (T_m / T - 1)$
6	liquids, solids	$\ln VP\ (atm) = - ([T_m - T\ ] / T)\ (8.5 - 5.0 \log \sigma + 2.3 \log \varphi) - (\ [T_b - T\ ] / T)\ (10 + 0.08 \log \varphi) + (\ [T_b - T\ ] / T - \ln [T_b / T]\ )\ (\ - 6 - 0.9 \log \varphi$

K_F = compound class-specific constant
R = gas constant (cal/mol K)
T = ambient temperature (K)
T_b = boiling point (K)
T_m = melting ppoint (K)
$T^* = T / T_b$
$T^{**} = T / T_m$
$C = 0.19\ T_b - 18$
m = specific factor for liquids (m=0.19) or solids (if T*>0.6, m=0.36; if 0.6> T* >0.5, m=0.8; if T* <0.5, m=1.19)
$m^* = - 0.2575\ T^{**} + 0.4133$
K = 0.803
MW = molecular weight
$f(T_m) = 2.9532\ (1 - T_m / T)$ for $T_m > T$; $f(T_m) = 0$ for $T_m < T$
$\ln \Delta p_{(s)}$ (equation 3) = term for solids only; $T_m / T - 1$(equation 5) = term for solids only
φ = conformational flexibility
σ = rotational symmetry number

Software: The estimation program MPBPVPWIN (SRC, 1997) uses for the estimation of the VP the modified Grain method, based on three separated methods for (estimated) boiling point. Estimates for volatile compounds (VP > 100 Pa) are significantly better than those of VP < 100 Pa.
Domain: (mainly non-ionic) organic substances. Not applicable for metals, metalloids and organometals.

4.1.5 Henry's Law constant (HE)

The partitioning of an organic chemical between water and air is described by the Henry's Law Constant. Chemicals with low values of HE will tend to partition into the aqueous phase at equilibrium. The HE is expressed either as the ratio of the partial pressure in the vapour phase and the concentration in water (Pa. m^3 mol^{-1}), or as the ratio of the concentrations in air and water (dimensionless). HE can be calculated from the ratio of the vapour pressure and the water solubility (not for miscible compounds and compounds with water solubility > 1 mol/l).
Software: HENRYWIN (SRC, 1997) based on fragment constants and interaction terms.
Domain: (mainly non-ionic) organic substances; logKow values between –3 and 7.

4.1.6 Abiotic degradation in the atmosphere

The atmospheric concentration of organic substances is determined by evaporation from terrestrial and aqueous compartments, removal by wet and dry deposition and transformations of the chemicals. The (indirect) photodegradation, i.e. reactions with reactive species formed by photochemical processes, has been recognised as the major transformation pathway for chemicals in the troposphere. The electrophilic addition of tropospheric radicals, hydroxyl radicals ($OH^{\cdot}$) and ozone (O_3) during day time, and $NO_3^{\cdot}$ radicals at night, constitutes the principal degradation pathway. In the EU report (1995) 3 models are described for $OH^{\cdot}$ or $NO_3^{\cdot}$ radical reaction rate constants and first vertical ionization energy as descriptor ($E_{i,v}$). These models are based on the Frontier Orbital Theory which postulates that the frontier molecular orbitals (HOMO and LUMO) are the most responsible for reactivity of chemicals (Sabljic & Güsten, 1990; Güsten et al., 1984).

1) $-\log k_{NO3} = 2.10 * E_{i,v} - 6.55$ (k=reaction rate constant) (n=62, r2=0.927, se=0.420)
2) $-\log k_{OH} = 0.79 * E_{i,v} + 3.06$ (n=129, r2=0.902, se=0.360)
3) $-\log k_{OH} = 1.52 * E_{i,v} - 2.06$ (n=32, r2=0.902, se=0.290)

The widely recommended method of estimation rate constants for the reaction of organic compounds with hydroxyl radicals has been derived by Atkinson (1987, 1988). The model is based on assumption that there are four possible reaction pathways in reaction of OH radicals with organic chemicals, that their reactions rate constants can be estimated separately, and that the estimation can be made by the additive group contribution approach. The four pathways include: H atom abstraction from C-H and O-H bonds ($k_{H\text{-}abst}$), addition of hydroxyl radicals to C-C multiple bonds ($k_{add.\ (C=C)}$). addition to aromatic ring ($k_{\ add./aromat.}$) and reactions with N,S, or P ($k_{\ N,S,P}$):

4) $k_{tot.} = k_{H\text{-}abst} + k_{add.\ (C=C)} + k_{\ add./aromat.} + k_{\ N,S,P}$

Three other models are described for $OH^{\cdot}$ reaction rate constants and the charge-limited effective HOMO-energy (ECH^C) and energy-limited LUMO-charge (QL^C) as descriptors (Klamt, 1993).

5) k_{OH}^C (add)= exp (1.29 * ECH_π^C (0.58) – 2.43 * QL^C (2.8) + 17.98)
(n=58, r2=0.954, se=0.280)
6) k_{OH}^C (ar) = exp (7.23 / (exp (5.352 * EEH^C (2.16) + 10.53)) – 0.165 * Δdef^C + 0.75)
(n=55, r2=0.970, se=0.340)
7) k_{OH}^C (abs) = exp (9.93 / (1 + exp (- 2.18 * (ECH^H (0.18) + 11.0)) – 8.03))
(n=46, r2=0.980, se=0.400)

(EEH^C = energy-limited effective HOMO-energy)
Δdef^C = energy required to deform the molecule in a way to enable the OH-addition)

Software: The group contribution model AOPWIN (SRC, 1997) is based on the method of Atkinson (see above) and estimates $OH^{\cdot}$ and ozone reaction rate constants using parameters for reaction centers and substituent constants.
Domain: alkanes, haloalkanes, alkenes, haloalkenes, polyenes, terpenes, alkynes, aldehydes, ketones, alcohols, glycols, ethers, esters, epoxides, thiols, thioethers, aliphatic amines, hydrazines, nitrites, nitrates, nitriles, P-containing organics, aromatic compounds (alkylbenzenes, halobenzenes, phenols, PAHs, styrene, methoxybenzene, aniline, nitro-benzene, biphenyl, dibenzofurans, dibenzodioxins).
Not applicable for perhalogenated alkanes; poor estimates for haloalkanes with 3 halogens on the same C atom, for small heterocyclic rings, for nitroalkanes, and for aromatic compounds which are not benzene derivatives.

4.1.7 Hydrolysis

The persistence of a chemical in the aquatic environment is, amongst others, dependent on the chemical reactions between the compound and water. Only a limited number of chemical classes is potentially hydrolysable. Rorije & Peijnenburg (EU, 1995) evaluated QSARs for abiotic degradation processes in the aqueous phase. From 31 models that described hydrolysis they selected (according to the selection criteria of Hermens et al. 1995) five models applying to brominated alkanes, esters, carbamates, and para-substituted benzonitriles (see flow-diagram 4.1).
Software: The HYDROWIN (SRC, 1997) program incorporates models for the estimation of the rate of hydrolysis of esters, carbamates, alkyl halides and epoxides in the environment. Models were not published and not evaluated in the EU-QSAR project.

Brominated alkanes : the model is based on linear regression and Taft's polar constant sigma (I) as descriptor.
1: *log k(i)/ k(o) = - 11.9 (3.5) x sigma (I)* (n=16, r2=0.77) k(o) = 4.09E-7 sec^{-1}
(k(i) = pseudo first order alkaline hydrolysis rate; k(o) = CH3-Br hydrolysis rate)
Range of descriptor values: sigma (I) from 0.1 to 0.3
Domain: log k(i)/k (o) from –5 to 1; saturated linear and branched bromoalkanes up to carbon atoms with bromo and/or phenyl substituents.

Esters: the model is based on multiple linear regression and Hammett and Taft constituent constants sigma, sigma*, and Es as descriptors.
For x-R-C(O))-R'-x' R,R'=alkyl or aryl
2: *log k = 0.98 (Es)R + 0.25 (Es)R' + 2.24(sigma*)R = 2.24(sigma*)R' + 2.09(sigma)x + 1.21(sigma)x' + 2.69* (n=103, r2=0.974)
k = second order alkaline hydrolysis rate constant, log k range from –5 to +5.
Domain: every chemical that contains ester bond –C(=O)-O-.

Carbamates: 2 models are selected that describe alkaline hydrolysis with rate constant k. Models are based on multiple linear regression and Hammett and Taft substituent constants sigma, sigma*, and Es as descriptors.
3: *logk = 2.39(sigma*)R1,R2 + 0.96(sigma)X1 + 7.97(sigma*)R3 + 2.81(sigma)X2 – 0.275*
(n=62, r2=0.973)
Domain: logk ranges from –6 to +6; all chemicals that contain a carbamate group with structure X1-R1-N(R2)-C(O)-O-R3-X3 where R2=hydrogen, R1 and R3 are either alkyl or arylgroups with X1 and X3 their respective substituents, and
4: *logk = 7.99(sigma*)R3 + 0.31(sigma)X2 + 3.14 (Es)R1,R2 + 0.442* (n=18, r2=0.903)

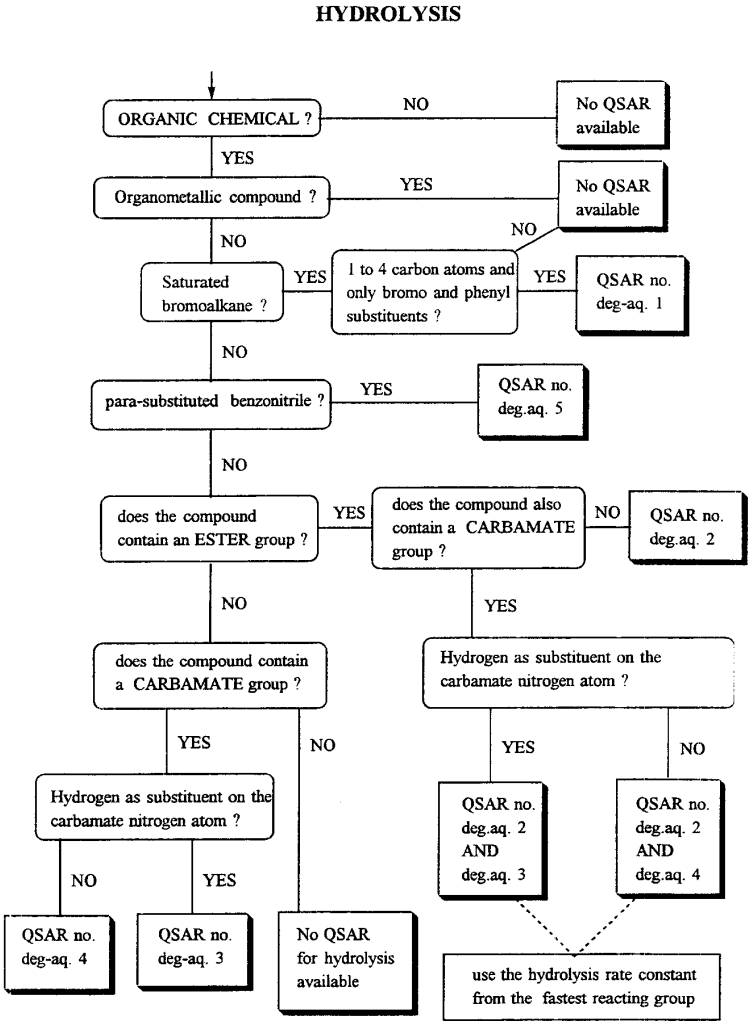


Figure 4 1: Flow-diagram for application of QSAR models for a number of potentially hydrolysable substances (Rorije en Peijnenburg, 1995). See text for degradation equations.

Domain: logk ranges from –7 to –2.5; all chemicals that contain a carbamate group with structure X1-R1-N(R2)-C(O)-O-R3-X3 where R2=alkyl or phenyl (**not** hydrogen), R1 and R3 are either alkyl or aryl with X1 and X3 their respective substituents.

Benzonitiles: model describes neutral and base-promoted hydrolysis with first order and pseudo first order rate constants; stepwise forward regression (least squares) with 7 initial descriptors and 1 final descriptor: Hammett sigma (para) constant
5: logk = 1.64 (±0.432) x sigma(para) – 1.37 (±0.170) (n=14, r2=0.858)

Domain: logk ranges from –2.7 to –0.2; all para-substituted benzonitriles.

4.1.8 Biodegradation

The persistence of chemicals in the environment is determined by the ability of ambient micro-organisms to utilise the chemicals as nutrient sources. This may reduce the concentration of the parent compound either through the formation of (potentially toxic) metabolites (primary degradation) or through the complete mineralisation of the substance to CO_2, water and mineral salts (ultimate degradation). Increasing biodegradability has been demonstrated in the presence of functional groups as carboxyl-, hydroxyl-, and methyl-groups and decreasing degradability in the presence of nitro-, amino-, cyano-, and halogen-groups. (Nenzda, 1998). Langenberg et al (1996) evaluated the use of QSARs for biodegradation in risk assessment. Numerous QSARs on biodegradation are reported of which 3 models are chosen as "best available":

- BIODEG database (1993) = BIOWIN (SRC, 1997): based on four models, a linear and a non-linear group contribution model and two expert judgement models, validated with experimental data obtained with the MITI-test. Classification in readily degradable (biodegradation probability = 1) and not readily degradable (biodegradation probability = 0). If the probability is close to 0.5 the result is suspect.

Domain: organic chemicals, not organometallic, with at least one of the substructures used as descriptor in the model.

- OECD#75: based on 7 fragment contributions. Classification in easily/non-easily degradable compounds (more/less than 60% degradation in 28 days)

Domain: acyclic aliphatic compounds, not a dithioether, a hydrazine or a compound with atoms other than C, N, H, O, P, S or halogen.

- OECD#78:based on 9 fragment contributions. Classification in easily/non-easily degradable compounds (more/less than 60% degradation in 28 days)

Domain: monocyclic aromatic compounds, substituents ONLY: -OH, -COO, -OCH3, -CH3, -SO2, -NO2, -NH2, -Halogen, -amido, non-terminal heteroatoms, branched alkyl chains. Mono AND polysubstitution is possible.

Hiromatsu et al.(2000) developed an empirical flowchart for predicting aerobic biodegradability of two groups of chemicals, monobenzene derivatives and acyclic compounds, based on the skeleton structure, or the interactions between the functional groups or the substructures bound to the skeleton structure. They catagorised the functional groups and the substructures bound of these two groups of chemicals into three fragment groups, negative, neutral, and negative. The negative and positive

groups are defined as those to inhibit or accelerate the biodegradation, respectively. The neutral group is defined as one that does not affect the biodegradability. The groups are listed in Table 4.3.
The negative group precedes the positive and neutral ones at trunk node, because it seems to have stronger effect on the biodegradation. When a chemical has a positive group without a negative, it tends to be readily biodegradable after considering combination of the other positive and neutral groups. The branch nodes (positive or negative factors) are basically judged from the number and the combination of the functional groups.

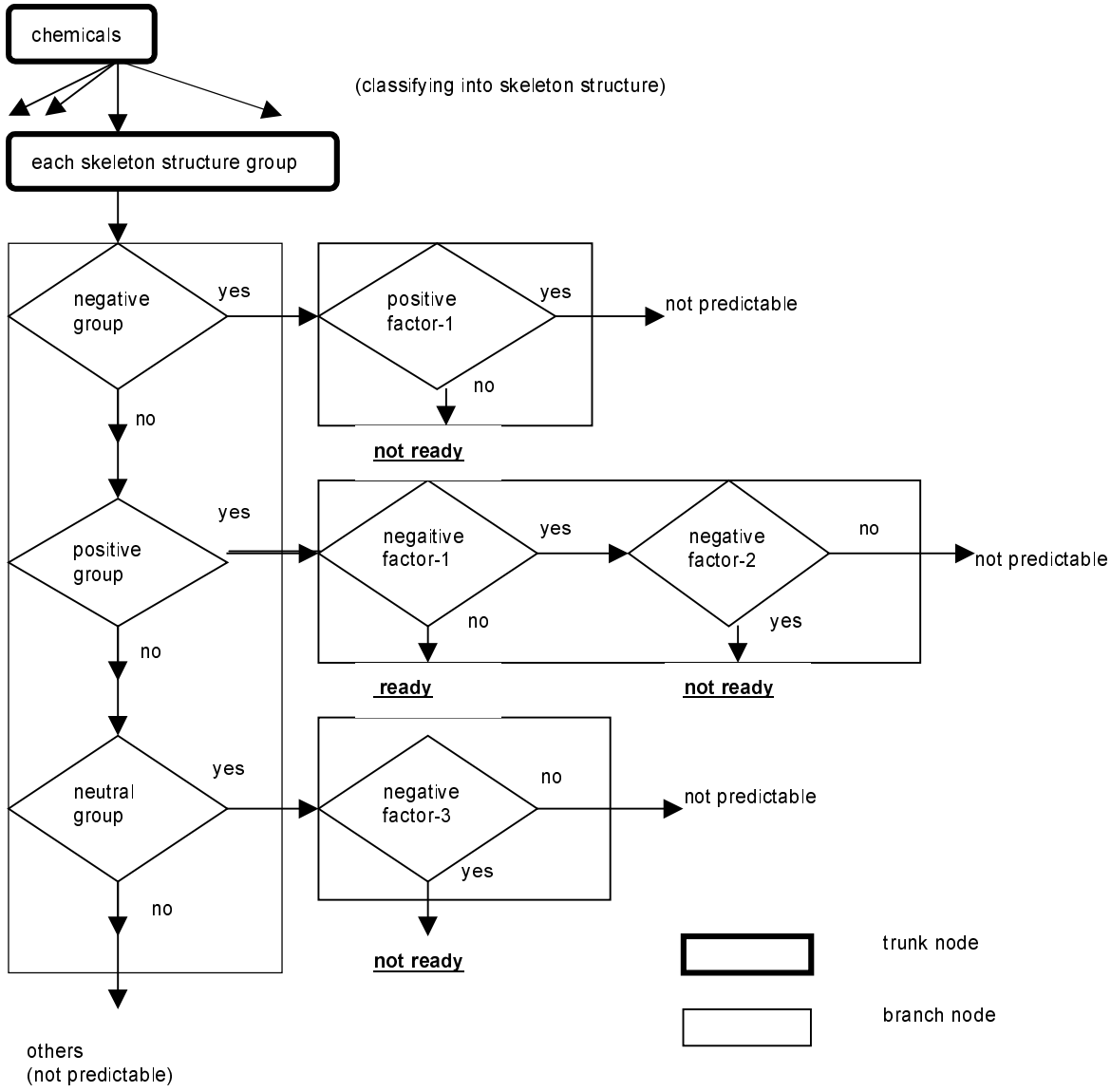


Figure 4.2: Predicting flowchart for biodegradability (Hiromatsu et al., 2000)

The flowchart was validated by using MITI data of 177 monobenzene derivatives and 168 acyclic compounds, resulting in correct prediction at 94% and 88% levels, respectively.

Table 4.3: Classification of groups based on effect on biodegradability defined from literature data (Hiromatsu et al., 2000)

	negative group	neutral group	positive group
mono benzene derivatives	NO_2, halogen, SO_3H, quaternary carbon, CF_3, tertiary amine, CN 2 substitution (*meta)* 3-6 substitution	CH_3, NH_2, OCH_3, ether, aldehyde and others 2 substitution (*ortho)* 2 substitution *(para)*	COOH, OH, ester, amide
acyclic compounds	NO_2, halogen, SO_3H, quaternary carbon, CF_3, plural substitution	CH_3, NH_2, OCH_3, ether, tertiary amine, CN and others	COOH, OH, ester, amide, aldehyde

4.1.9 Bioconcentration (BCF)

BCF for aquatic organisms

The bioconcentration factor (BCF) is defined as the ratio between the concentration of the chemical in biota and the concentration in water at equilibrium. The bioconcentration factor can also be calculated by the ratio of the first order uptake and elimination rate constants, a method that does not require equilibrium conditions.

For calculation of the BCF the TGD (1996) suggest 2 models for substances with logKow < 6 and > 6, respectively. It is generally agreed that a linear relationship exist for chemicals which are not biotransformed with logKow < 6. Many examples of such models have been published in the literature (Veith et al. (1979, 1980), Könemann & van Leeuwen (1980), MacKay (1982), Nenzda (1991, Bintein et al.(1993). Linear model developed by Veith et al. (1979):

- log BCF = 0.85 logKow – 0.70 (n=55, r2=0.90).

In case of substances with logKow > 6 linear modelling of bioconcentration is inaccurate as in this range the measured logBCF data tend to decrease with increasing logKow. Explanations for non-linearity mainly refer to either biotransformation, reduced membrane permeation kinetics or reduced biotic lipid solubility for large molecules. A model in which the influence of non-equilibrium conditions has been eliminated is developed by Connell & Hawker (1988) as a polynominal equation:

- $\log BCF = 6.9 \times 10^{-3} (\log Kow)^4 - 1.85 \times 10^{-1} (\log Kow)^3 + 1.55 (\log Kow)^2 - 4.18 \log Kow + 4.79$ (n=45, r2=nd).

Because the statistical validity of the polynomial relationship is questionable the model has been recalculated, resulting in a significant parabolic relationship:

- $\log BCF = - 0.20 (\log Kow)^2 + 2.74 \log Kow - 4.72$ (logKow range of 6-10) (n=43, r2=0.78).

In the protocol for derivation of harmonised Maximum Permissible concentrations (MPCs) by Kalf et al. (1999) the following equation is proposed:

- BCF = Corganism/C water or k1/k2 (rate constants for uptake and depuration)

- If measured values are not available the BCF can be predicted from the relationship between Kow and BCF:
- BCF fish (whole body, fresh weight) = 0.05 * Kow and BCF mussel (whole body, fresh weight) = 0.013 * Kow.

Nenzda (1991) has established a non-linear relationship between BCF and Kow for broader ranges of logKow(1.0-11.2):

- log BCF = 0.99 logKow – 1.47 * log (4.97 * 10^{-8}* Kow * +1) + 0.0135 (n=132).

Lu et al. (1999) established a BCF estimation model by means of molecular connectivity indices (MCI):

- logBCF = 0.757 $^{0}\chi^{v}$ – 2.650 $^{1}\chi^{v}$ + 3.372 $^{2}\chi$ – 1.186 $^{2}\chi^{v}$ – 1.807 $^{3}\chi_{c}$ + 0.770 (n=80, r2=0.907, s=0.364)

Software: the program BCFWIN (SRC, 1996) estimates the BCF from the molecular structure of the compound. BCFWIN recognises substances that are likely to be biotransformed, which result in decreased BCF values (Sijm et al., 1999).
Domain: Above mentioned QSAR equations are applicable for neutral, non-polar, no-ionic organic chemicals. Not for ionizable or charged molecules, and not for metals and metalloids.

BCF for terrestrial organisms - earthworms
For the assessment of secondary poisoning in the terrestrial food chain the bioconcentration factor in worms is necessary:

- BCF worm = Cworm/C (pore)water or BCF worm = C worm . Koc / C soil(oc)

Connell & Markwell (1990) developed a QSAR equation for earthworms based on the data of van Gestel & Ma (1988) and other BCF values in the literature:

- log BCF = 1.0 logKow – 0.6 (n=100, r2=0.91).

Jager (1998) proposed a BCF based on the theoretical relation :

- BCF = Fwater + F fat * Kow (Fwater: volume fraction water in worm; Ffat: volume fraction fat in worm)

which resulted in:

- BCF = 0.84 + 0.012*Kow

Domain: 1 < logKow < 6, neutral organic, non ionic substances, not organometallics.

4.1.10 Soil sorption (normalised to organic carbon)

The sorption to soil components is a determinant factor for the mobility of chemicals, accounting for their distribution among soil, sediment and water phases. The extent to which chemicals partition between the solid and solution phases determines the likelihood of the chemicals leaching through the soil to groundwater or being immobile. The extent of sorption to soils is governed by a variety of physicochemical properties of both the soil and the chemical. Relevant soil parameters include: organic carbon content, size of particles, porosity, clay content, biological activity and biomass, humidity, pH, CEC, and temperature.
Sorption occurs if the free energy of interaction between the soil components and the chemical is negative. Sorption may be caused by the following processes: van der Waals interactions, hydrophobic bonding, hydrogen bonding, charge-transfer interactions, ligand exchange and ion binding, direct and induced ion-dipole and dipole-dipole interactions and covalent binding. For estimation of the adsorption coefficient several models are available (see flow-diagram 4.3).

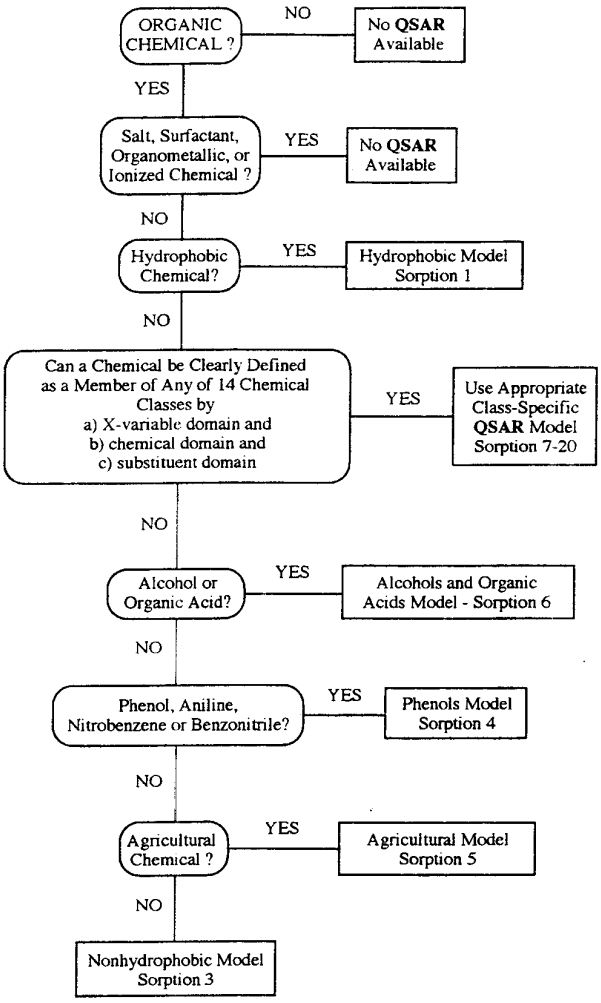


Figure 4.3: Flow-diagram for application of QSAR models for estimating adsorption coefficients (Güsten & Sabljic, 1995). See text for equations.

Based on first-order molecular connectivity index (CHI 1)

1. log Koc = 0.52 * CHI 1 + 0.70 g/l (CHI 1 can be calculated by GRAPHIII program)
(n=81, r2=0.962, se=0.264)

range of descriptor values : CHI 1: 1 to 11; compounds with 3 to 22 halogen atoms
domain: all chemicals that contain only C,H,F,Cl, Br, and I atoms.

2. log Koc = 0.81 * logKow + 0.10 g/l (n=81, r2=0.888, se=0.451)
range of descriptor values: 1 < logKow < 7.5
domain: all chemical that contain only C, H, and halogen (F, BR, Cl, I) atoms.

3. log Koc = 0.52 * logKow + 1.02 g/l (n=390, r2=0.632, se=o.557)
range of descriptor values: - 2 < logKow < 8
domain: all hydrophobic chemical classes, except salts, surfactants, and organometalics.

4. log Koc = 0.63 * logKow + 0.90 g/l (n=54, r2=0.749, se=0.401)
range of descriptor values: 1 < logKow < 5
domain: substituted phenols (Cl, Br, CH_3, OH. NO_2, CH_3O)
substituted anilines (Cl, Br, CH_3, CF_3, CH_3O, N-Me)
chlorinated benzonitriles
substituted nitrobenzenes (Cl, Br, NH_2).

5. log Koc = 0.47 * logKow + 1.09 g/l (n=216, r2=0.682, se=0.425)
range of descriptor values: - 1 < logKow < 8
domain: agricultural chemicals (acetanilides, carbamates, esters, phenylureas, phosphates, triazines, triazoles, and uracils).

6. log Koc = 0.47 * logKow + 0.50 g/l (n=36, r2=0.773, se=0.388)
range of descriptor values: - 1 < logKow < 5
domain: alcohols (alkyl, phenalkyl, OH substituents) and all types of organic acids.

models 7 – 20: Class specific soil sorption

acetanilides:	log Koc = 0.40 * logKow + 1.12 g/l (logKow: 0.9 to 5) (n=21, r2=0.513, se=0.339)
alcohols:	log Koc = 0.39 * logKow + 0.50 g/l (logKow: - 1 to 5) (n=13, r2=0.768, se=0.397)
amides:	log Koc = 0.33 * logKow + 1.25 g/l (logKow: - 1 to 4) (n=28, r2=0.461, se=0.491)
anilines:	log Koc = 0.62 * logKow + 0.85 g/l (logKow: 1 to 5) (n=20, r2=0.818, se=0.341)
carbamates:	log Koc = 0.365 * logKow + 1.14 g/l (logKow: - 1 to 5) (n=43, r2=0.578, se=0.408)
dinitroanilines:	log Koc = 0.38 * logKow + 1.92 g/l (logKow: 0.5 to 5.5) (n=22, r2=0.826, se=0.242)
esters:	log Koc = 0.49 * logKow + 1.05 g/l (logKow: 1 to 8) (n=25, r2=0.763, se=0463)
nitrobenzenes:	log Koc = 0.77 * logKow + 0.55 g/l (logKow: 1 to 4.5) (n=10, r2=0.703, se=0.583)
organic acids:	log Koc = 0.60 * logKow + 0.32 g/l (log Kow: - 0.5 to 4) (n=23, r2=0.748, se=0.336)
phenols & benzonitriles:	log Koc = 0.57 * logKow + 1.08 g/l (logKow: 0.5 to 5.5) (n=24, r2=0.748, se=0.373)
phenylureas:	log Koc = 0.49 * logKow + 1.05 g/l (logKow: 0.5 to 4.5) (n=52, r2=0.624, se=0.335)
phosphates:	log Koc = 0.49 * logKow + 1.17 g/l (logKow: 0 to 6.5) (n=41, r2=0.733, se=0.452)
triazines:	log Koc = 0.30 * logKow + 1.50 g/l (logKow: 1.5 to 4) (n=16, r2=0.332, se=0.379)
triazoles:	log Koc = 0.47 * logKow + 1.405 g/l (logKow: - 1 to 5) (n=15, r2=0.657, se=0.482)

Software: the estimation program PCKOCWIN (SRC, 1996) based on fragment distributions.

All the above mentioned software programs of SRC are avialable in one program: EPIWIN.
An example of an EPIWIN estimation record is given in Appendix IV.

4.2 QSARs for environmental effect assessment

4.2.1 Classification of (organic) chemicals

One critical factor in QSAR-based effect assessment is the proper assignment of a chemical to a mode of action and associated QSAR: a solid classification on the mode of action would facilitate the implementation of QSARs in effect assessments. Therefore specific attention is paid to the results of various classification studies in the last decade.

Verhaar et al. (1992) defined a number of distinctive classes of chemicals that could either be assigned a mode of action, or that could otherwise be assigned a quantitative relationship between the structure of those chemicals falling in it and their acute aquatic toxicity. Four classes of chemicals were defined, and each of these classes is assigned a QSAR derived method to enable the estimation of the acute aquatic toxicities:

1) *Inert chemicals*
mode of action in aquatic toxicity is narcosis, an absolutely nonspecific mode of action. Potency of a chemical to induce narcosis is entirely dependent on its hydrophobicity. Narcosis type toxicity is also called "baseline" toxicity or minimum toxicity. Chemicals can be modelled using logKow as sole descriptor.

2) *Less inert chemicals*
slightly more toxic than baseline toxicity, not reactive chemicals when considering overall acute effects. Compounds acting by "polar narcosis" mechanism. Without elaborating the distinction between nonpolar and polar narcosis, it can still be stated that it is possible to identify structural requirements for these chemicals. These chemicals are usually characterised as possessing hydrogen bond donor acidity (e.g. phenols and anilines).

3) *Reactive chemicals*
can have all kind of different modes of action and have enhanced toxicity compared to baseline toxicity. These chemicals react unselectively with certain chemical structures commonly found in biomolecules (e.g. epoxides, which react with sulfhydryl groups of cysteine residues of peptides), or chemicals that are metabolised into more toxic species (bioactivation).

4) *Specifically acting chemicals*
form a diverse set of chemicals that exhibit toxicity due to specific interaction with certain receptor molecules, such as acethylcholinesterase inhibitors (organic phosphorus esters) or chemicals which react with sodium channel regulating receptors in neurons (specific or receptor toxicity).

In order to classify a certain compound as belonging to one of the above-mentioned classes a classification scheme was proposed (see Appendix I). Compounds that cannot classified as belonging to class 1, 2 or 3 and that are not known to be compounds acting by a specific mechanism (4) can only be classified as "not possible to classify according to these rules".
In general, effect concentrations for less inert chemicals (class 2) are between 5 and 10 times lower than predicted by baseline toxicity QSAR equation, whereas effect concentrations for reactive chemicals (class 3), as well as for specifically acting chemicals (class 4) are between 10 and 10^4 times lower than predicted by baseline toxicity QSAR equation. These observations can be used to define toxicity range factors RF_Ts, multiplication factors to estimate baseline toxicity of a chemical, as

presented in figure 4.4 These RF_Ts define a range of effect concentrations that will encompass the true effect of this chemical.
In order to show that these RF_Ts constitute an acceptable, albeit rather broad estimation of effect concentrations a rationalisation is given in which Toxic Ratio's (TR= LC50, baseline/LC50, experimental) for a number of chemicals classified are compared with the RF_T for the class that these chemicals belong to. From the result is was clearly demonstrated that narcotics and polar narcotics are two distinct classes of chemicals with narrow distribution curves. Reactive and specific chemicals, on the other hand, do not significantly differ from each other; moreover, the wide curve for class 4 suggest this class be an inhomogeneous class.

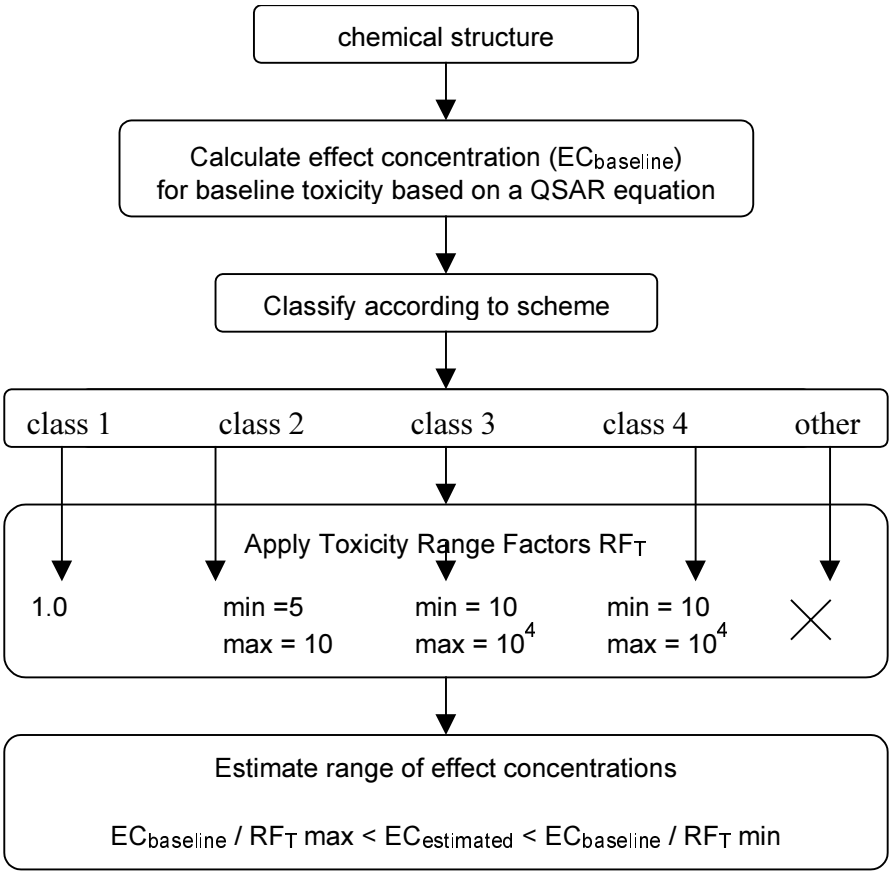


Figure 4.4: Outline of the classification/QSAR approach (Verhaar et al., 1992)

The rules of this system rely on the presence or absence of certain structural or substructural features to assign a compound to one of the four classes. Such a rule-based system has the disadvantage that it is impossible to classify objects that do not fit the existing rules, even though they might easily fit the effect the classification is based upon. Therefore, it would be advantageous to have a classification system that can separate compounds into different classes based on a set of molecular descriptors. Verhaar et al. (1996) generated a set of mainly quantum chemical molecular

descriptors for a large training set of compounds for which the classification into either class 1 or class 2 is known unambiguously. In total there were 172 different compounds in the combined training set, namely 50 baseline toxicity compounds and 122 polar narcosis compounds including mainly phenols, anilines and mononitrobenzenes. The difference between non-polar and polar narcosis might be linked to the polar nature of the class 2 compounds and more specifically to the capability of forming hydrogen bonds as either acceptor or donor molecules (or possibly both). Therefore, molecular descriptors chosen included quantitative indicators of the hydrogen bond donor and acceptor capacities of the molecule. In view of the polar nature of polar narcotic compounds quantum chemical parameters that model the polarity or polarizability of a molecule were also selected. Additionally, two parameters loosely connected with reactivity and three parameters concerning hydrophobic interactions were selected. The resultant training data set was subsequently subjected to PLS discriminant analysis in order to investigate the usefulness of the chosen descriptors for separating polar and non-polar narcosis compounds. The resultant training data set was also subjected to PLS and multiple linear regression analysis to see whether the descriptors could in principle be used to create a single, large aquatic toxicity QSAR for these two groups of compounds.
The results of both the discriminant analysis and the aquatic toxicity QSAR model showed that hydrogen bond acceptor, and less important hydrogen bond donor capacities are the prime determinants in describing the difference between class 1 and 2 compounds.

Sixt & Altschuh (1997) investigated whether the classification scheme of Verhaar et al. (1992) based on fish toxicity data is applicable to photobacteria data (Microtox, 30 min log 1/EC50). They classified for this purpose 240 diverse chemicals according to this scheme into 3 classes: nonpolar narcotics (82), polar narcotics (81) and reactive chemicals (77). A simple plot of the Microtox data with logKow for all Class 1 compounds showed a rather poor correlation. However, it was found that for some compounds, for which toxicity was overestimated, the EC50 values were much higher than their experimental water solubilities. These compounds (20) were removed from the plot as well as 23 compounds lying considerably above the line to derive a baseline quantitative structure-activity relation for photobacterium toxicity:

log 1/EC50 = 1.07 * logKow - 2.36 (r2=0.95, n=45, sd=0.32).

In comparison with fish baseline toxicity Microtox baseline toxicity increases more rapidly with logKow, but is lower at small values of logKow. Using the simple statistics described by Verhaar et al. (1992) the distributions of the relative toxicity $\log T_R$ values for Class 1 and 2 were rather broad, indicating that not all these compounds were baseline toxicants.
Many of the Class 1 compounds were considerably more toxic than the baseline model predicts. For Class 2 and 3, the mean values of $\log T_R$ were close together. It seemed that Class 2 and 3 are mixed to some extent, i.e., some compounds classified as polar narcotics are reactive and vice versa. From these results, it had to be stated that the rules given by Verhaar et al. (1992) cannot be used for the Microtox data without modification.
Despite this, rather good QSARs based on quantum chemical descriptors could be established for the 3 different Classes (see chapter 4.2.2: Table 1, 2 and 3).

Clements et al. (1995) of the US EPA examined the distribution of 8234 discrete organic chemicals by chemical class and found that 10 chemical classes account for 97.6% of the total with the remainder of 2.4% being distributed on another 30 classes. The dominant class was formed by the neutral organics with 44.7% of the chemicals. followed by esters with 17.4%, acids 12.2%, amines 8.6%, phenols 5.6%, aldhydes 3.1%, anilines 3.0%, acrylates 1.4%, diketones 0.8%, thiols 0.8%, other classes 2.4%. (see also ECOSAR).
Utilising the same database of 8234 chemicals an evaluation was made of the distribution of 563 EU's High Production Volume Chemicals (HPVCs) by chemical class. Overall, a general agreement was found in the distribution with some exceptions. Neutral organics formed the dominant class with 51.2% of the chemicals examined and there was a considerable drop in the representation of esters (9.2%).

Schultz et al. (1997) identified the mechanisms of toxic action of 198 **phenols** to *Tetrahymena pyriformis* from molecular descriptors. All chemicals were *a priori* assigned a mode of action (MOA) using the following rules based on molecular structure.
Weak acid respiratory uncouplers:
- more than 1 nitro group, or
- more than 1 cyano group or
- more than 3 halogen groups, or
- single nitro group and more than 1 halogen group

Soft electrophiles:
- single nitro goup but not more than 1 halogen group

Precursor to soft electrophiles (pro-electrophile):
- 2 or more hydroxy groups in the ortho or para position and at least 1 unsubstituted aromatic carbon atom, or
- amino group in the ortho or para position to the hydroxy group and at least 1 unsubstituted aromatic carbon atom

Precursor to redox cyclers (pro-redox cycler):
- 2 or more hydroxy groups in the ortho or para position and no unsubstituted carbon atom, or
- amino group in ortho of para position to the hydroxy group and no unsubstituted aromatic carbon atom

Polar narcosis:
- molecule did not meet any of the above structural criteria.

A previous study (Cronin & Schultz, 1996) in which a variety of physico-chemical descriptors (43) were calculated, demonstrated that partitioning and electrophilicity (E_{LUMO}) are important to govern the toxicity of phenols. LogKow values were corrected for the effect of ionisation because it was demonstrated in previous studies that ionisation has a profound effect on phenol toxicity. The toxicity of the weak acid respiratory uncouplers, soft electrophiles, pro-electrophiles, and pro-redox cyclers are all in excess of the toxicity of polar narcotics. Based on E_{LUMO} there was a separation of phenols according to MOA (Mode Of Action). Phenols exhibit bioreactive toxicity. Bioreactivity can be subdivided into noncovalent reversible or covalent irreversible mechanisms. Noncovalent bioreactive mechanisms of action include polar narcosis, weak acid respiratory uncoupling, and redox cycling oxidative stress. These mechanisms perturb biological systems without heavy atom changes. Covalent-

mediated bioreactivity results in chemical change in biological systems and is specific for different biochemical mechanisms (i.e. Schiff-base formation, Michael-type acceptor, nucleophilic substitution). Soft electrophilicity and pro-electrophilicity represent covalent-mediated toxic mechanisms.
In this study it was clearly demonstrated that chemicals in the same class can exhibit several MOAs.

Bearden & Schultz (1998) evaluated the toxicity data of 256 chemicals tested in both the 96-h *Pimephales promelas* mortality assay (LC50) and the 2-d *Tetrahymena pyriformis* growth inhibition assay (IGC50) using QSARs. The study demonstrated that with specific mechanism-based exceptions the *Tetrahymena* and *Pimephales* assays can be used to predict the toxicity of the other. Each chemical was *a priori* assigned a toxic mode of action of either narcoses or soft electrophilicity. This separation is based on the presence or absence of specific structural features which theoretically correspond to mechanism of action. Narcoses were separated into nonpolar narcosis, polar narcosis, monoester narcosis, diester narcosis, amine narcosis, and weak acid respiratory uncoupling. QSARs for each narcotic-type mechanism were initially developed based solely on hydrophobicity quantitated as logKow. The slopes of these logKow based QSARs were observed to ascertain whether a relationship exists between the value of the slope and the reactivity of the mechanism of action. It was shown that as relative reactivity increased, the slope value of the logKow QSAR decreased.
The soft electrophile mode of action was separated into the specific molecular mechanisms of $S_N 2$ reactors, Schiff-base formers, Michael-type addition or proelectrophilicity (precursors to Michael-type addition chemicals). These mechanisms were represented structurally by the nitrobenzenes, aldehydes, polarised α-β unsaturates (e.g. acrylates and methacrylates), and acetylenic alcohols, respectively. Electrophilic toxicity was not correlated with hydrophobicity. QSARs based on molecular orbital quantum chemical descriptors were used to improve the predictability of the electrophilic mechanisms: average superdelocalizability for the nucleophilic addition of nitrobenzene, atom *x* and *y* acceptor superdelocalizability and bond order for the Michael-type addition of the acrylates, and logKow and atom *x* net charge for the Schiff-forming aldehydes. The pertinent descriptors for proelectrophiles were logKow and average superdelocalizability. Principal differences between QSARs for the two biological endpoints were observed for the ester narcoses, proelectrophiles, and Schiff-base forming aldehydes.
Statistical comparison between *Tetrahymena* and *Pimephales* for all 256 chemicals resulted in the following model:

log 1/LC50 = 0.975 x 1/IGCC50 + 0.573 (r2=0.736; q2=0.731; s=0.633)

Weyers et al. (2000) examined the relative sensitivity of fish, *Daphnia*, and algae and the correlation of their EC50/LC50 values by comparing their acute toxicity data from the New Chemicals Database of the European Chemicals Bureau. Correlation of EC50/LC50 values against physico-chemical properties (logKow and water solubility) was also investigated. The results could be relevant to possible exemption from full testing requirements, choice of test organisms, and revision of notification

requirements. The algal growth inhibition test was found to be the most sensitive, giving the lowest value in 43.5% of the cases, in 18.6% of the cases no test was most sensitive, and in 37.9% of all cases, either fish or *Daphnia* tests were more sensitive. The best correlation of the EC50/LC50 values was found for fish versus *Daphnia* with $r^2 = 0.597$. The plot of *Daphnia* versus algae, and fish versus algae showed a much lower r^2 values of 0.348 and 0.366, respectively. The correlations between logKow and the toxicity values were lower than between toxicity values.

Russom et al.(1997) stated that development of an expert system to predict acute mode of toxic action from chemical structure first requires a knowledge base from which rules can be derived. This knowledge base was derived from analysis of 617 chemicals in the fathead minnow acute toxicity database (Brooke, Geiger et al., 1984, 1985, 1986, 1988). Chemicals from the fathead minnow database were evaluated through analysis of a variety of endpoints: [a] dose-response relationships in the 96 h bioassay, [b] behavioural responses in the 96 h LC50 bioassays, [c] excess toxicity values (predicted LC50/observed LC50), [d] results of joint toxic action studies with fathead minnows, [e] fish acute toxicity syndrome (FATS) studies on rainbow trout, [f] examination of toxicological literature specific to the issue on toxicodynamic classification. Using this data set, a computer based expert system has been established whereby chemical structures are associated with likely modes of toxic actions and, when available, corresponding QSARs. Each chemical was classified into one of eight modes of action:

- base-line (nonpolar) narcosis or narcosis I
- polar narcosis or narcosis II
- ester narcosis or narcosis III
- oxidative phosphorylation uncoupling
- respiratory inhibition
- electrophile/proelectrophile reactivity
- acetylcholinesterase inhibitors
- CNS seizure responses

and presented the following results:

Polar narcotics and uncouplers of oxidative phosphorylation:
- anilines
- phenols
- pyridines

Ester narcosis:
- esters
- diesters

Acetylcholinesterase inhibitors:
- carbamates
- organophosphorous compounds

Neurotoxicants:
- organochlorines (cyclodiend and DDT type)
- pyrethroids

Reactive chemicals:
- acylation based reactivity
- isocyanate based reactivity
- carbonyl based reactivity
- alkylation and arylation based reactivity
- sulfhydriyl based reactivity
- nitroso based reactivity
- nitrile based reactivity
- oxime based reactivity
- beta-halogenated alcohol reactivity
- n-halogenated acetophenone reactivity

Other:
- dinitrobenzenes
- pyridium compound
- quaternary ammonium compounds
- chloro-diesters
- acetamidophenols
- quinoline

Basak et al. (1998) used molecular similarity, neural networks, and discriminant analysis methods to predict acute toxic modes of action for a set of 283 chemicals. The majority of these chemicals had been previously determined through toxicodynamics studies in fish to be narcotics (class I and 2), electrophiles/proelectrophiles, uncouplers of oxidative phosphorylation, acethylcholinesterase inhibitors, and neurotoxicants. Nonempirical parameters, such as topological indices and atom pairs, were used as structural descriptors for the development of similarity-based, statistical, and neural network models. Rates of correct classification ranged from 65 to 95% for the 283 chemicals.

These studies indicate that a general classification of organic chemical substances is feasible (Verhaar et al., 1992), but that the rules derived for one species not necessarily can be applied to other species without modification (Sixt & Altschuh, 1997) or only are applicable after an in-depth classification (Bearden & Schultz, 1998). Although most of the organic chemicals are likely to be neutral organics (Clements et al., 1995) belonging to narcotics (class 1 and 2), the use of new descriptors learns that classification into narcotics is not without problems (Basak et al., 1998) and even a wide array of narcotics can be distinguished (Bearden & Schultz, 1998).

4.2.2 Chemical class related QSARs for effect assessment

QSARs for environmental effect are presented in Table 4.4, 4.5, 4.6, and 4.7 for nonpolar narcotics, polar narcotics, reactive chemicals, and specific acting chemicals, respectively.

Table 4.4: QSARs for non-polar narcotic chemicals

Species	Endpoint	Equation + Statistics	Domain	Reference
pisces				
Pimephales promelas	96 h LC50 (mol/l)	log LC50 = - 0.85 logKow – 1.39 n=58, r2=0.94 Q2=0.93; se=0.36	– 1.24 < logKow< 5.13	Veith et al., 1993 (EU project + TGD)
Pimephales promelas	96 h LC50 (mol/l)	log LC50 = - 0.94 logKow + 0.94 log(6.8 x 10^{-5}Kow + 1) – 1.25 n=60, r2=0.97, se=0.35	logKow: – 1.2 – 5.2	Veith et al., 1983
Pimephales promelas	96 h LC50 (mmol/l)	log 1/LC50 = 0.79 logKow – 1.35 n=47, r2=0.921, se=0.40	logKow: 0 – 5	Nenzda & Russom, 1991
Pimephales promelas	8 d LC50 (mol/l)	log 1/LC50 = Σ (T) + 3.34 n=105, r2=0.921, se 0.31		Hall et al., 1989
Cyprinus carpio	24 h LC50 (mol/l)	log 1/LC50 = 0.919 LogKow + 0.967 n=5, r2=0.985, se=0.151	logKow: – 0.7- 1.4	Hansch & Dunn, 1972
Carassius auratus	24 h LC50 (mol/l)	log 1/LC50 = 0.881 LogKow + 0.989 n=5, r2=0.958, se=0.250	logKow : – 0.7 – 1.4	Hansch & Dunn, 1972
Carassius auratus	24 h LC50 (mmol/l)	log 1/lc50 = 1.0 logKow – 2.2 n=8, r2=0.98, se=0.18	logKow: 0.6 – 3.1	Lipnick et al., 1987
Cyprinodon variegatus	96 h LC50 (mmol/l)	log 1/LC50 = 0.50 $^{o}\chi^{v}$ – 1.90 n=19, r2=0.92, se=0.37	$^{o}\chi^{v}$: 2.7 – 8.6	Sabljic, 1983
Brachydanio rerio *Pimephales promelas*	28-32 d NOEC ELS test (mol/l)	log NOEC = - 0.90 logKow – 2.30 n=27, r2=0.92, Q2=0.91, se=0.33	0.46 < logKow < 5.18	van Leeuwen et al, 1990 (EU proj.+TGD)
Brachydanio rerio	28 d NOEC (µmol/l)	log 1/NOEC = 1.06 (±0.09) logKow – 4.57 n=6, r2=0.987, se=0.17	chlorobenzenes	van Leeuwen et al., 1990
Poecilia reticulata	7/14 d LC50 (mol/l)	log LC50 = - 0.869 logKow – 1.193 n=50, r2=0.960, Q2=0.957, se=0.314	- 1.4 < logKow < 5.18	Könemann, 1981 (EU project)
Poecilia reticulata	EC50 (mol/l)	log EC50 = 0.687 * $(V_{ee})^{(N)}$ + 4.309 n=36; r2=0.889; se=0.241; q2=0.877	benzene derivatives	Girones et al., 1999
crustacea				
Daphnia magna	48 h EC50 (mol/l)	log EC50 = - 0.94 logKow – 1.32 n=49, r2=0.95, Q=0.94, se=0.34	- 1.36 < logKow < 5.98	Abernethy et al. (1996) (EU, TGD) (a)
Daphnia pulex	96 h LC50 (mol/l)	log LC50 = - 0.91 logKow – 0.99 n=18, r2=0.9, Q2= 0.88, se=0.37	0.88 < logKow < 5.18	Ikemoto et al., 1992 (EU project)
Daphnia pulex	96 h LC50	log LC50 = - 1.0 logKow – 1.70	2.13 < logKow < 5.67	Govers et al., 1984

Species	Endpoint	Equation + Statistics	Domain	Reference
	(mol/l)	n=5, r2=0.995, Q2=0.98, se=0.13	PAHs (just containing C and H)	(EU project)
Daphnia magna	16 d NOEC (mol/l)	log NOEC = - 1.05 logKow – 1.85 n=10, r2=0.97, Q2=0.95, se=0.39	- 0.24 < logKow < 5.18	Hermens et al., 1984 (Eu proj + TGD) (b)
Nitroca spinipes	NOEC (mol/l)	log NOEC = - 0.78 LogKow – 2.14 n=16, r2=0.95, se=0.39	logKow: – 0.77 to 5.13	van Leeuwen et al., 1992
Daphnia magna	18-21 d NOEC reprod.(mol/l)	log NOEC = - 1.04 LogKow – 1.70 n=17, r2=0.98, se=0.25	logKow: – 0.24 to 5.18	van Leeuwen et al., 1992
algae				
Ankistrodesmus falcatus	4 h EC50 (mol/l)	log EC50 = - 0.90 logKow – 0.65 n=13, r2=0.97, Q2=0.96; se=0.13	2.13 < logKow < 5.18	Wong et al., 1984) (EU proj.)
Chlorella vulgaris	3 h EC50 (mol/l)	log EC50 = - 0.95 logKow – 0.34 n=34, r2= 0.92, Q2=0.91, se=0.32	0.88 < logKow < 5.98	Ikemoto et al., 199 (EU proj.) (c)
Pseudokirchneriella subcapitata	72-96 h EC50 growth (mol/l)	logEC50 = - 1.00 logKow – 1.23 n=10, r2=0.93, Q2=n.d., se=0.17	logKow : 1 to 6	van Leeuwen et al., 1992 (TGD)
Pseudokirchneriella subcapitata	3 h EC50 (mol/l)	log EC50 = - 1.18 logKow – 0.15 n=6, r2=0.97, Q2=0.92, se= 0.22	2.89 < logKow < 5.73	Calamari et al., 1983 (EU proj.)
Skeletonema costatum	96 h NOEC (mol/l)	log NOEC = - 0.72 LogKow – 1.42 n=9, r2=0.72, se=0.45	logKow: 1.48 to 4.60	van Leeuwen et al., 1992
Scenedesmus subspicatus	48 h NOEC (mol/l)	log NOEC = - 0.86 LogKow – 1.41 n=8, r2=0.83, se=0.44	logKow: 0.76 to 3.53	van Leeuwen et al., 1992
Pseudokirchneriella subcapitata	72/96 h NOEC (mol/l)	log NOEC = - 1.00 LogKow – 1.71 n=10. r2=0.95, se=0.17	logKow :2.19 to 4.05	van Leeuwen et al., 1992
bacteriophyta				
Vibrio fisheri	30 min EC50 (mmol/l)	log 1/EC50 = 1.07 * logKow – 2.36 n=45, r2=0.95, s=0.32		Sixt & Altschuh, 1997
Vibrio fisheri	30 min EC50 (mmol/l)	log 1/EC50 = 0.29 α_m – 0.24E_{LUMO} – 3.15 n=67, r2=0.88, s=0.50		Sixt & Altschuh, 1997
Vibrio fisheri	30 min EC50 (mmol/l)	log 1/EC50 = 0.35 α_m – 0.19 E_{LUMO} + 1.36 I_{anis} – 4.46 n=66, r2=0.91, s=0.42		Sixt & Altschuh, 1997
Vibrio fisheri	15 min EC50 (unit ?)	log 1/EC50 = 0.99 logKow – 2.08 n=6, r2=0.988, s=0.24	alkanones	Cronin & Schultz, 1998
Bacillus subtilis	IC50 (mol/l)	log IC50 = 0.882 * $(V_{ee})^{(N)}$ + 2.191 n=19; r2=0.863; s=0.363; q2=0.828	benzene derivatives	Girones et al., 1999
Clostridium botulinum	24 h NOEC	log NOEC = - 0.82 LogKow – 0.29	logKow 0.77-6.11	van Leeuwen et al., 1992

Species	Endpoint	Equation + Statistics	Domain	Reference
	(mol/l)	n=14, r2=0.94, se=0.46		
Baccillus subtilis	30 min NOEC (mol/l)	log NOEC = - 0.64 LogKow – 2.03 n=14, r2=0.92, se=0.33	logKow - 0.77-4.57	van Leeuwen et al., 1992
Psudomonas putida	6 h NOEC growth (mol/l)	log NOEC = - 0.64 LogKow – 1.60 n=5, r2=0.98, se=0.14	logKow – 0.25-2.72	van Leeuwen et al., 1992
Vibrio fisheri	15 min. NOEC (mol/l)	log NOEC = - 0.68 LogKow – 1.52 n=20, r2=0.90, se=0.60	logKow - 1.31- 4.14	van Leeuwen et al., 1992
Pseudomonas fluorescens	20 min EC50 (mg/l)	log EC50 = - 0.90 + 0.33 logS n=11, r2=0.94	chlorobenzenes	Boyd et al., 1998
Pseudomonas fluorescens	20 min EC50 (mg/l)	log EC50 = 0.59 – 1.20 logKow n=11, r2=0.76	chlorobenzenes	Boyd et al., 1998
Pseudomonas fluorescens	20 min EC50 (mg/l)	log EC50 = 5.06 + 2.04 logH n=11, r2=0.72	chlorobenzenes	Boyd et al., 1998
Pseudomonas fluorescens	20 min EC50 (mg/l)	log EC50 = 5.47- 0.33 SA n=11, r2=0.70	chlorobenzenes	Boyd et al., 1998
Pseudomonas fluorescens	20 min EC50 (mg/l)	log EC50 = 5.24 – 0.41MV n=11, r2=0.70	chlorobenzenes	Boyd et al., 1998
activated sludge	IC50 (mmol/l)	log IC50 = 3.43 – 0.75 $^{1}\chi^{v}$ n=14, r2=0.832	alcohols, ketones, esters	Xu & Nirmalakhandan, 1998
activated sludge	IC50 (mmol/l)	log IC50 = 2.02 – 0.63 $^{1}\chi^{v}$ n=11, r2=0.750	alkanes	Xu & Nirmalakhandan, 1998
activated sludge	IC50 (mmol/l)	log IC50 = 2.92 – 0.48 $^{o}\chi^{v}$ n=14, r2=0.785	halogenated aliphatics	Xu & Nirmalakhandan, 1998
insecta				
Chironomus riparius	48 h LC50 (mol/l)	log LC50 = - 1.02 * log Kow – 0.72 n=9, r^2 = 0.993,		Roghair et al., 1996
Aedes egypti	48 h NOEC mort. (mol/l)	log NOEC = - 1.09 LogKow – 1.36 n=14, r2=0.96, se=0.27	logKow: – 1.36 to 2.72	van Leeuwen et al., 1992
Culex pipiens	48 h NOEC mort.(mol/l)	log NOEC = - 0.86 LogKow – 1.98 n=5, r2=0.95, se=0.33	logKow: – 0.25 to 2.72	van Leeuwen et al., 1992
Sigara striata	7-day NOEC mort.(mol/l)	log NOEC = - 9.12 logKow – 1.52 n=9, r2=0.996	logKow: -0.77 to 5.18	Evers & Henzen, 1992
annelida				
Branchiura sowerby	72 h LC50 (mol/l)	log LC50 = - 1.06 * logKow – 0.39 n=8,r^2 = 0.980		Roghair et al., 1996
fungi				
Saccharomyces cerevisiae	24 h NOEC	log NOEC = - 0.78 LogKow – 0.35	logKow: - 0.77 to1.56	van Leeuwen et al., 1992

Species	Endpoint	Equation + Statistics	Domain	Reference
	(mol/l)	n=5, r2=0.90, se=0.29		
protozoa				
Tetrahymena pyriformis	48 h NOEC (mol/l)	log NOEC = - 0.80 LogKow – 1.28 n=26, r2=0.93, se=0.40	logKow: - 0.77 to 5.58	van Leeuwen et al., 1992
coelenterata				
Hydra oligactis	48 h NOEC mort. (mol/l)	log NOEC = - 0.86 LogKow – 2.05 n=5, r2=0.92, se=0.45	logKow: – 0.25 to 2.72	van Leeuwen et al., 1992
mollusca				
Lymnea stagnalis	48 h NOEC mort. (mol/l)	log NOEC = - 0.86 LogKow – 2.08 n=5, r2=0.96, se=0.30	logKow: – 0.25 to 2.72	van Leeuwen et al., 1992
amphibia				
Ambystoma mexicanum	48 h NOEC mort. (mol/l)	log NOEC = - 0.88 LogKow – 1.89 n=5, r2=0.94, se=0.36	logKow: – 0.25 to 2.72	van Leeuwen et al., 1992
Rana temporaria	30 min NOEC mort. (mol/l)	log NOEC = - 1.09 LogKow – 1.47 n=11, r2=0.98, se=0.23	logKow: – 0.77 to 2.97	van Leeuwen et al., 1992
Xenopus Laevis	48 h NOEC mort.(mol/l)	log NOEC = - 0.90 LogKow – 1.79 n=5, r2=0.94, se=0.38	logKow: – 0.25 to 2.72	van Leeuwen et al., 1992

T: fragment constants for substituents on monoaromatic compounds (T_{NO2} = 0.58 etc...)
α_m: molecular polarizability
$^1\chi^v$ = first order valence molecular connectivity index
I_{anis}: anisotropy index
$^o\chi^v$ = zero order valence connectivity index
E_{HOMO}: energy of the highest occupied molecular orbital
logS: log aqueous solubility
logH: log Henry's Law constant
V_{ee}: molecular quantum electron repulsion energy
E_{LUMO}: energy of the lowest unoccupied molecular orbital
SA: surface area
MV: molecular volume

a: see also Bobra et al. (1983) Hermens et al (1984); Ikemoto et al. (1992)
b: see also de Wolf et al. (1988)
c: see also Bobra et al (1983)

Table 4.5: QSARs for polar narcosis (excess toxicity to non-polar narcosis)

Species	Endpoint	Equation + Statistics	Domain	Reference
algae				
Chlorella pyrenoidosa	72 h EC50 (mol/l)	log EC50 = - 0.51 logKow – 0.06E_{HOMO} – 1.23 Q^- + 0.23 E_{LUMO} – 1.67 Q^+ n=11; r2=0.89	phenols, nitrobenzenes, anilines	Urrestarazu et al., 1999
Chlorella pyrenoidosa	72 h NOEC (mol/l)	log EC50 = - 0.68 logKow – 0.05 E_{HOMO} – 1.49 Q^- + 0.39 E_{LUMO} – 2.31 Q^+ n=11; r2=0.89	phenols, nitrobenzenes, anilines	Urrestarazu et al., 1999
pisces				
Pimephales promelas	96 h LC50 (mol/l)	log LC50 = - 0.73 logKow – 2.16 n=86, r2=0.90, Q2=0.90, s.e.=0.33	- 1.31 < logKow < 6.21	Cajina-Quezada et al., 1990 (EU + TGD) (e)
Pimephales promelas	96 h LC50 (mol/l)	log 1/LC50 = 0.80 logKow – 1.80 n=23, r2=0.96, se=0.27	aliphatic and aromatic amines	Schultz et al., 1991
Brachydanio rerio *Pimephales promelas*	28-32 d NOEC (mol/l)	log NOEC = - 0.64 logKow – 4.01 n=4, r2=0.96, Q2=0.73, s.e.=0.21	0.90 < logKow < 3.69 anilines	van Leeuwen et al., 1990 (EU proj.)
Poecilia reticulata	14 d LC50 (mol/l)	log LC50 = - 0.64 logKow – 2.54 n=56, r2=0.82, Q2=0.81; s.e.=0.25	0.90 < logKow < 4.28	Benoit-Guod et al., 1984 (EU proj.) (d)
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 0.988 logKow – 4.02 n=17, r2=0.8855, se=0.39	anilines chloro/alkyl anilines	Hermens et al., 1984
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 0.749 $\Sigma\Pi^-$ - 2.90 n=17, r2=0.906, se=0.51	anilines chloro/alkyl anilines	Hermens et al., 1984
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 0.359 $\Sigma\Pi^-$ + 745 $\Sigma\sigma$ –2.73 n=17, r2=0.937, se=0.30	anilines chloro/alkyl anlines	Hermens et al., 1984
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 1.29 $\Sigma\sigma$ –2.50 n=17, r2=0.917, se=0.33	anilines chloro/alkyl anlines	Hermens et al., 1984
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 0.41 logKow + 0.85 $\Sigma\sigma$ – 3.17 n=17, r2=0.935, se=0.31	anilines chloro/alkyl anilines	Hermens et al., 1984
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 0.922 logKow – 3.72 n=11, r2=0.946, se=0.27	anilines chloroanilines	Hermens et al., 1984
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 0.245 logKow + 1.17 $\Sigma\sigma$ – 3.04 n=11, r2=0.970, se=0.22	anilines chloroanilines	Hermens et al., 1984
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 0.701 $\Sigma\Pi^-$ - 2.74 n=11, r2=0.928, se=0.31	anilines chloroanilines	Hermens et al., 1984

Species	Endpoint	Equation + Statistics	Domain	Reference
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 0.200 $\Sigma\Pi^-$ + 1.17 Σσ –2.80 n=11, r2=0.973, se=0.20	anilines chloroanlines	Hermens et al., 1984
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 1.56 Σσ –2.76 n=11, r2=0.967, se=0.21	anilines chloroanlines	Hermens et al., 1984
Poecilia reticulata	LC50 (mmol/l)	log 1/LC50 = 0.63 logKow + 2.52 n=17, r2=0.982, s=0.16	phenols	Saarikoski, & Vluksela, 1982
fish	LC50 (mmol/l)	log LC50 = - 0.62 logKow + 0.40 n=78, r2=0.86	chlorinated phenols	Clements & Nabholtz, 1994
fish	LC50 (mmol/l)	log LC50 = - 0.64 logKow + 0.72 n=20, r2=0.82	aliphatic amines	Nabholtz & Platz, 1990
crustacea				
Daphnia magna	24/48 h EC50 (mol/l)	log EC50 = - 0.56 logKow – 2.79 n=37, r2=0.77, Q2=0.73, s.e.=0.37	- 0.86 < logKow < 4.01	Devillers & Chambon, 1986 (EU proj. + TGD) (f)
Daphnia magna	EC50 (mmol/l)	log EC50 = - 0.58 logKow – 0.52 n=10, r2=0.78	aliphatic amines logKow =< 6	Nabholtz & Platz, 1990
Daphnia magna	EC50 (mol/l) pH 6.0	log 1/EC50 = 0.631 logKow + 2.95 n=23, r2=0.756 se=0.365	anilines + phenols	Cronin et al., 2000
Daphnia magna	EC50 (mol/l) pH 6.0	log 1/EC50 = 0.338 logKow + 3.65 n=13, r2=0.810 se=0.138	anilines	Cronin et al., 2000
Daphnia magna	EC50 (mol/l) pH 6.0	log 1/EC50 = 0.964 logKow + 1.93 n=9, r2=0.990 se=0.112	phenols	Cronin et al., 2000
Daphnia magna	EC50 (mol/l) pH 7.8	log 1/EC50 = 0.437 logKow – 0.0868 pKa – 9.17 q_H + 5.68 n=21, r2=0.785 se=0.250	anilines + phenols	Cronin et al., 2000
Daphnia magna	EC50 (mol/l) pH 7.8	log 1/EC50 = 0.278 logKow + 3.91 n=13, r2=0.741 se=0.137	anilines	Cronin et al., 2000
Daphnia magna	EC50 (mol/l) pH 7.8	log 1/EC50 = 0.822 logKow + 2.03 n=9, r2=0.961 se=0.196	phenols	Cronin et al., 2000
Daphnia magna	EC50 (mol/l) pH 9.0	log 1/EC50 = 0.324 logKow – 0.104 pKa – 10.4 q_H + 6.34 n=24, r2=0.814 se=0.247	anilines + phenols	Cronin et al., 2000
Daphnia magna	EC50 (mol/l) pH 9.0	log 1/EC50 = 0.537 logKow + 2.55 n=9, r2=0.889, se=0.224	phenols	Cronin et al., 2000
bacteriophyta				
Vibrio fisheri	30 min EC50 (mmol/l)	log 1/EC50 = 0.61 logKow + 4.1 q^+ - 15.6 q_{mean} + 2.0 q_H + 0.35 E_{LUMO} – 0.27 n=73, r2=0.72, se=0.46		Sixt & Altschuh, 1997

Species	Endpoint	Equation + Statistics	Domain	Reference
Vibrio fisheri	30 min EC50 (mmol/l)	log 1/EC50 = 0.489 logKow + 0.126 n=16. r2=0.848, se=0.267	phenols, anilines	Schultz & Cronin, 1997
protozoa				
Tetrahymena pyriformis	2- d ICG50 (mmol/l)	log 1/IGC50 = 0.31 logKow – 0.74 n=17, r2=0.927, se=0.10	saturated aliphatic mono-carboxylic acids	Seward & Schultz, 1999
Tetrahymena pyriformis	2- d ICG50 (mmol/l)	log 1/IGC50 = 0.19 logKow – 0.66 n=9, r2=0.951, se=0.08	saturated alophatic di-carboxylic acids	Seward & Schultz, 1999
Tetrahymena pyriformis	2- d ICG50 (mmol/l)	log 1/IGC50 = 0.26 logKow – 0.21 E_{LUMO} – 0.46 n=38, r2=0.727, se=0.219	all aliphatic carboxylic acids	Seward & Schultz, 1999
Tetrahymena pyriformis	48-h ICG50 (mol/l)	log 1/ICG50 = 0.72 logKow – 1.64 n=20, r2=0.92, se=0.29	aliphatic and aromatic amines	Schultz et al., 1991
amphibia				
Rana japonica	24-h LC50 (mol/l)	LC50 = 0.92 logKow – 0.41 pKa + 3.32 n=19, r2=0.90, se=0.20	phenols	Wang et al, 2000

d: see also Deneer at al. (1987; Hermens et al. (1984); Könemann & Musch (1981); Roberts (1986 a & b); Saarikoski & Viluksela (1982)
e: see also Gombar (1986); Hall & Kier (1984); Newsome et al. (1991); Schultz et al.(1986); Veith & Broderius (1986)
f: see also Nenzda & Klein (1987); Vighi & Calamari (1987)

E_{LUMO}: energy of the lowest unoccupied molecular orbital
E_{HOMO}: energy of the highest occupied molecular orbital
Q+ most positive partial charge on a hydrogen atom
Q-: most negative partial charge on a non-hydrogen atom
q_H: sum of atomic charges on hydrogen
q_O: sum of atomic charges on oxygen
q_{mean}: mean atomic charge
Π and Hammet σ constituent constants

Table 4.6: QSARs for reactive chemicals

Species	Endpoint	Equation + Statistics	Domain	Reference
pisces				
Poecilia reticulata	14 d LC50 (µmol/l)	- log LC50 = (0.39 ± 0.05) logKow + (3 ± 0.4) log k_1 – 2.25 n=14,	epoxy compounds	Deneer et al., 1988a
Poecilia reticulata	14- d LC50 (µmol/l	- log LC50 = (0.36 ± 0.04) logKow – 2.54 n=14, r2=0.923, s=0.19	aldehydes	Deneer et al., 1988b
Poecilia reticulata	14- d LC50 (µmol/l	- log LC50 =(0.36 ± 0.04) logKow – (0.08 ± 0.05) log k_1 – 2.32 n=14, r2=0.938, s=0.18	aldehydes	Deneer et al., 1988b
Poecilia reticulata	14-d LC50 (µmol/l)	log LC50 = - 0.33 logKow – 1.02 logk_{NBP} – 1.51 n=15, r2=0.81	small organochlorine electrophiles*	Verhaar et al., 1996
bacteriophyta				
Vibrio fisheri	30 min EC50 (mmol/l)	log 1/EC50 = 0.17 α_m + 6.9 RSD_{max} – 1.6 q_H – 3.85; n=70, r2=0.62, s=0.67		Sixt & Altschuh, 1997
Vibrio fisheri	15 min EC50 (?)	log 1/EC50 = 0.55 logKow – 0.58 n=6, r2=0.994, s=0.07	aldehydes (Schiff base-forming)	Cronin & Schultz, 1998
Vibrio fisheri	15 min EC50 (?)	log 1/EC50 = 0.52 logKow + 0.35 n=6, r2=0.914, s=0.19	aldehydes (Michael-type addition)	Cronin & Schultz, 1998
Vibrio fisheri	15 min EC50 (?)	log 1/EC50 = 0.79 logKow - 1.17 E_{LUMO} – 0.41 n=19, r2=0.892, s=0.46	ketones, aldehydes	Cronin & Schultz, 1998
crustacea				
Daphnia magna	48 h EC50 (mol/l)	logEC50 = – 1.367 + 0.737E – 2.28E^2 – 8.645sq^- n=16, r2=0.910, s=0.16	substituted benzaldehydes	Jin et al., 1998

E = sum of Hammett σ values of the substituent groups attached to the carbon of benzene cycle
sq^- = the second most negative net atomic charge on a atom
RSD_{max}: maximum atomic radical superdelocalizability
k_1: second order rate constant for reaction with L-cysteine
k_{NBP}: first order rate constant for alkylation of 4-nitrobenzylpyridine
E_{LUMO}: energy of the lowest unoccupied molecular orbital
q_H: sum of atomic charges on hydrogen
α_m: molecular polarizability

small organochlorine electrophiles:

2,2'-dichlorodiethylether	chloroacetone	allylchloride
2,3-dichloro-1-propene	propargylchloride	4-nitrobenzylbromide
3,4-dichloro-1-butene	α,α'-dichloro-m-xylene	α-chlorotoluene
3-chloro-1-butene	t-1,3-dichloropropene	c-1,3-dichloropropene
t-3-chloro-2-butene	t-1,4-dichloro-2-butene	2,4,α-trichlorotoluene

Table 4.7: QSARs for specifically acting chemicals

Species	Endpoint	Equation + Statistics		Reference
Daphnia magna	EC50 (µmol/l)	log 1/EC50 = 0.55 logKow – 0.085 $(logKow)^2$ – 3.931 – 0.016 $(^1\chi^v ox)^2$ + 0.13 α_3 – 0.0006 $(\alpha_3)^2$ + 4.58 n=22, r2=0.895, s=0.67	organophosphorous pesticides	Vighi et al.,1991
pisces				
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = 0.23 Σπ + 0.80 log k_{NBP} + 2.77 n=9, r2=0.92, se=0.19	AChE inhibitors log k_{NBP} – 4-	Hermens et al., 1987
Poecilia reticulata	14 d LC50 (µmol/l)	log 1/LC50 = - 0.41 logKow + 31.4 *dq* (PO) – 48.0 n=9, r2=0.93,	AChE inhibitors logKow 2.5 –5.5 *dq*(PO) 1.44 –1.47	Schuurmann, 1990
Pimephales promelas	96 h LC50 (mol/l)	log LC50 = - 0.59 logKow – 3.22 n=6, r2=0.91	uncouplers	Schultz et al., 1986
Pimephales promelas	96 h LC50 (mmol/l)	log LC50 = - 0.67 logKow + 0.05 n=11, r2=0.91	uncouplers	OECD, 1992
Insecta				
Chironomus riparius pH 6, 11°C	24-h EC50 (µmol/l)	EC50 = 0.91(0.22) – 0.129(0.023) $^1\chi^v$ + 0.077(0.02) $^3\kappa$ n=10, r2=0.84	organophosphorous insecticides	Landrum et al., 1999
Chironomus riparius pH 6, 18°C	24-h EC50 (µmol/l)	EC50 = 0.96(0.22) – 0.128(0.023) $^1\chi^v$ + 0.058(0.02) $^3\kappa$ n=10, r2=0.83	organophosphorous insecticides	Landrum et al., 1999
Chironomus riparius pH 6, 25°C	24-h EC50 (µmol/l)	EC50 = 0.51(0.10) – 0.067(0.01) $^1\chi^v$ + 0.028(0.01) $^3\kappa$ n=10, r2=0.86	organophosphorous insecticides	Landrum et al., 1999
Chironomus riparius pH 7, 11°C	24-h EC50 (µmol/l)	EC50 = 1.28(0.23) – 0.18(0.025) $^1\chi^v$ + 0.092(0.024) $^3\kappa$ n=10, r2=0.89	organophosphorous insecticides	Landrum et al., 1999
Chironomus riparius pH 7, 18°C	24-h EC50 (µmol/l)	EC50 = 0.57(0.11) – 0.079(0.012) $^1\chi^v$ + 0.039(0.01) $^3\kappa$ n=10, r2=0.87	organophosphorous insecticides	Landrum et al., 1999
Chironomus riparius pH 7, 25°C	24-h EC50 (µmol/l)	EC50 = 0.20(0.09) – 0.031(0.009) $^1\chi^v$ + 0.024(0.009) $^3\kappa$	organophosphorous insecticides	Landrum et al., 1999

Species	Endpoint	Equation + Statistics		Reference
		n=10, r2=0.84		
Chironomus riparius pH 8, 11°C	24-h EC50 (µmol/l)	EC50 = 0.57(0.20) – 0.086(0.02) $^{1}\chi^{v}$ + 0.064(0.02) $^{3}\kappa$ n=10, r2=0.75	organophosphorous insecticides	Landrum et al., 1999
Chironomus riparius pH 8, 18°C	24-h EC50 (µmol/l)	EC50 = 0.94(0.20) – 0.12(0.02) $^{1}\chi^{v}$ + 0.050(0.02) $^{3}\kappa$ n=10, r2=0.88	organophosphorous insecticides	Landrum et al., 1999
Chironomus riparius pH 8, 25°C	24-h EC50 (µmol/l)	EC50 = 0.65(0.10) – 0.085(0.011) $^{1}\chi^{v}$ + 0.031(0.01) $^{3}\kappa$ n=10, r2=0.90	organophosphorous insecticides	Landrum et al., 1999

log k_{NBP}: 1st order rate constant for alkylation of 4-nitrobenzylpyridine
dq(PO)*:* charge density difference between the phosphorous and the oxygen of the phenoxy leaving group
$^{1}\chi^{v}$ = first order valance molecular connectivity index
$^{3}\kappa$: third order of simple shape parameters

4.3 Computerised QSAR models for effect assessment

The following computerised QSAR models have been developed:

- ECOSAR (Ecological Structure Activity Relationships) is a personal computer software program that is used to estimate the toxicity of chemicals used in industry and discharged into water. The program is developed by US EPA at Washington and predicts the toxicity of chemicals to aquatic organisms such as fish, invertebrates, and algae by using Structure Activity Relationships (SARs) based on logKow as sole descriptor. The program estimates acute (short-time) toxicity and, when available, chronic (long-term or delayed) toxicity. ECOSAR allows access to over 100 SARs developed for 42 chemical classes. The SARs contained within the program are based on experimental data and many of them have been validated. ECOSAR can be searched with a CASnumber or with a description of the chemical structure using Simplified Molecular Input Line Entry System (SMILES). ECOSAR uses the same underlying Smilescas databank as is used in the software program EPIWIN, so both programs can be used along with each other. ECOSAR designates with an asterix when the estimated L(E)C50 or chronic value ChV for a chemical is above the solubility of the substance. In addition the upper limits are given for the logKow above which values the prediction for different species are not longer reliable. ECOSAR is a very user-friendly program, however, it has to be noted that ECOSAR estimates for a number of chemicals toxicity data on the basis of only two, or in some cases even one, experimental data. An example of a record from ECOSAR is given in Appendix III.
- Probabilistic Neural Network (PNN): chemical descriptors required for input in the PNN methodology include the log MW and the presence/absence of 33 molecular descriptors (e.g., -CN, -COOH, -O, =O, SO_x, etc), the number of C, H, As, Br, Cl, F, Fe, Hg, I, K, Mn, N, Na, O, P, S, Se, Si, Sn, and Zn atoms present in the molecule, and the logarithm of the ratio of the MW corresponding to all atoms present over the total MW. The PNN methodology used for generating predictions a training dataset of 865 compounds from the TerraTox toxicity database.
- Computational Neural Network (CNN): eight input descriptors are required in the CNN methodology which are entered into an 8-6-1 (8 descriptors, 6 neurons, 1 output) three layer, fully feed-forward neural network to generate acute toxicity predictions for fathead minnows. The CNN methodology used to develop the predictions was originally based on a training set of 375 organic compounds from the COMPUTOX toxicity database (Eldred et al, 1999). Toxicity predictions cannot be performed for chemicals with a triply-bonded O, or S, or those containing P atoms, have a triazine ring, sulfoxide moieties or quaternary N (Kaiser and Niculescu, 1999).
- ASTER (ASsessment Tools for the Evaluation of Risks): QSAR database developed by US EPA at Duluth using chemical structure fragments to predict likely modes of toxic action. ASTER can be searched with a CASnumber or with a description of the chemical structure using Simplified Molecular Input Line Entry System (SMILES). ASTER includes models to estimate toxicity for narcosis I, polar narcosis, ester narcosis, and oxidative phosphorylation uncoupling modes of action (Russom et al, 1997) All QSARs in ASTER are based on logKow as sole descriptor of toxicity and estimates are only provided inside the limits of the

QSAR models. The expert system within ASTER is based on a training set of 617 organic chemicals. The QSARs are validated by integration with the AQUIRE database for toxic effects.

- OASIS uses different modelling approaches for nonspecifically acting chemicals and reactive chemicals. For development of the QSAR models for nonspecifically acting chemicals the Duluth fish toxicity database (Brooke, Call, Geiger, 1982-1990) was employed. QSAR modelling of these 217 narcotics chemicals, with the exception of amines and esters, which were modelled separately, showed logKow (calculated by ClogP) and the electrophilicity descriptor E_{LUMO} as the most important descriptors. In OASIS quantum-chemical descriptors associated with reactive groups and logKow were used to model the toxicity of reactive chemicals: maximum donor delocalizability of benzene ring atoms for phenols and anilines; maximum donor delocalizability and acceptor delocalizability at α-C-atom for α,β-unsaturated alcohols; charge at carbonyl oxygen for aldehydes; bond order between carbon and halogens for α,β-unsaturated halides, etc.
- The UBA(FGH)-SAR-System (IUCT, 1992) provides tools for estimating physico-chemical (logKow, vapour pressure, solubility, boiling point, Henry's coefficient, Koc and pKa), toxicological (BCF, aquatic toxicity, mammalian toxicity and mutagenicity) and degradation properties (atmospheric degradation and biodegradation). The procedures are based on molecular structure and correlations with measured properties of chemicals with similar structure as well as on property-property relationships.

Recently a comparison was made between the first six above mentioned QSAR modelling packages by generating predictions for acute (96-h) toxicity to the fathead minnow (*Pimephales promelas*). A number of summary statistics were calculated to enable comparison of model performance. The PNN model was found to be the best, all other packages showed their own disadvantages and limitations. However, to use PNN model knowledge of such systems and specific training is required. In generally ECOSAR performed better than ASTER and CNN. (QSAR workshop, Canada, 2000).

Other software for QSAR modelling for ecotoxicology purposes are:

- M-CASE: In 1994 Klopman and co-workers developed a computerised program based on human toxicological endpoints (see Hulzebos et al. (1999) p.23) The same multiple Computer-Automated Structure Evaluation Program (M-CASE) was used by Klopman et al (1999) for the construction of an acute fish toxicity model based on a wide series of experimental data for guppies. The M-CASE program is a QSAR expert system which is based on a hierarchical algorithm designed for the treatment of large databases (learning sets) consisting of structurally different compounds. The fundamental assumption of the M-CASE methodology is that if a substructure is not relevant for the observed activity, it will be found randomly in both active and inactive compounds. If it is related to the observed activity, it will be found predominantly in active and marginally active compounds. This substructure is called a biophore. The M-CASE program starts off by identifying the statistically most significant substructure that exists within the learning set. The molecules containing this biophore are removed from the database and the remaining molecules are submitted to a new analysis for the

next biophore (or toxicophore). This procedure is repeated until either the activity of all the molecules in the learning set have been accounted for or no more statistically significant substructures can be found. In this way, M-CASE logically breaks the database into subsets of molecules, each associated with a particular biophore.The M-CASE program has recently been expanded with a new feature called the baseline activity identification algorithm (BAIA), which allows it to identify baseline activity due to a specific physical attribute (e.g.Kow) of the molecule. All chemicals whose toxicity is not accounted for by BAIA are assumed to derive their activity from other factors, which can then be identified by M-CASE analysis. Thus, for the purpose of M-CASE analysis, activity is defined as the residual activity obtained by subtracting the relevant baseline activity from the observed one.

- CoMFA: Briens et al (1995) studied the applicability of comparative molecular field analysis (CoMFA) in ecotoxicology. CoMFA was used to relate (by PLS) ecotoxicological data with steric (van der Waals) and electrostatic (Coulombic with 1/r dielectric) fields of chemicals forming an intentionally selected homogeneous set, in this case the chlorophenols. From the available data 16 different biological systems were selected, leading to predictive CoMFA QSARs in 14 of the cases. This was attested by cross-validation and bootstrapping, which also authorised the prediction of the chlorophenols toxicity values, when they are missing. The established relations proved to have an excellent predictive power.
- AMSOL: Bureau et al., (1997) studied the applicability of the free energies of solvation for the prediction of ecotoxicity. Free energies of solvation of chlorophenols were calculated in two solvents, water and n-hexadecane from the AMSOL program. These free energies of solvation are the sum of two terms: the first accounts for the electric polarization of and by the solvent; the second component accounts for the free energy necessary to form a cavity in the solvent to make room for the solute and for the changes in dispersion interaction and solvent structure that accompany the solvation process. Regression analysis was used to correlate the free energies of solvation to the Microtox toxicity data of 19 chlorophenols and PLS was used to correlate the free energies of solvation to several biological data (inhibition of bacterial growth, inhibition of phenol degradation, toxicity for *Daphnia magna* and *Brachydanio rerio*). An excellent relation between the descriptors and the ecotoxicity data was found.
- CATALYST: Briens et al., (1999) used CATALYST, a 3D-QSAR based software system, originally developed for pharmaceutical research, to model ecotoxicity data. CATALYST is a new approach that focuses on modelling chemical-receptor interaction from the point of view of the receptor, using information derived only from the chemical. Molecular structures are treated as templates consisting of strategically positioned chemical functions that will bind effectively with complementary functions on receptors. Toxicity data for chlorophenols were used for the formation of a training set, consisting of approximately 20 molecules ranging in activity over four orders of magnitude. The training set is used to derive a hypothesis, a minimal collection of chemical features common across the set that explain the observed activity. The resulting model can be used to predict the activity of molecules not included in the training set, by comparing how well the chemical features of a subject molecule overlap with the chemical features in the model (hypothesis). Examination of standard errors and r2 values of estimations revealed high predicting power.

A comparison was made between the above described CATALYST and CoMFA models studied by Briens et al. (1995, 1999), the free energies of solvation model studied by Bureau et al.(1997) and a model with logKow as sole descriptor from Ribo & Kaiser (1983). Statistics (correlation coefficients and standard errors of estimations) of QSARs obtained with the four different models were obtained for 5 data sets which were common in the four studies. The results are presented in Table 4.8.

Table 4.8: Correlation coefficients and standard errors of estimation of QSAR obtained with four methods

	CATALYST		CoMFA		G_{CDSWH}		logKow	
	r^2	se	r^2	se	r^2	se	r^2	se
A (n=18)	0.857	0.310	0.950	0.190	0.843	0.327	0.719	0.438
B (n=19	0.910	0.213	0.950	0.167	0.875	0.248	0.752	0.349
C (n=19	0.908	0.087	0.882	0.104	0.794	0.133	0.834	0.119
D (n=18)	0.832	0.268	0.896	0.159	0.878	0.166	0.777	0.224
E (n=18)	0.848	0.193	0.879	0.178	0.852	0.191	0.734	0.256

A: *Bacillus*
B: *Photobacterium phosphoreum*
C: inhibition of phenol degradation
D: *Daphnia magna*
E: *Brachydanio rerio*

Examination of the standard errors (*se*) revealed that the CoMFA model showed the best predictive power, with the lowest *se* values in four cases, and significant differences compared to the *se* values of the other models. The next most powerful model involves the CATALYST hypothesis, with *se* values lower than those obtained with the G_{CDSWH} descriptor in three of the five studies.
Examination of the correlation coefficients (r^2) showed that the CATALYST and the CoMFA model possess the best prediction power. For the chlorophenols training set, and for these 5 data sets, the classical approach, consisting of the logKow value as descriptor, is found to be less accurate.

5. Recommendations for use of QSARs within RIVM/CSR

5.1 Starting points

From the viewpoint of supply (science) and demand (policy) the following statements can be derived from the foregoing chapters:

- Classification of chemicals according to their mode of action and the derivation of reliable QSARs is very complex theoretically. These days there is an array of potentially useful descriptors, models and statistical techniques of which the new approaches seem to be most promising but still are in a developmental stage, and when becoming available, will require a high level of expertise for implementation in risk assessments. Only a limited number of QSARs for a few processes and/or chemical classes is directly available for use, mostly related to exposure assessments.
- The policy needs for using QSARs in environmental hazard and risk assessment by RIVM/CSR seem to be limited. In most frameworks experimental toxicity data will be available and preferred. Only for existing substances more effort may be required in the QSAR approach, especially with respect to intervention values for sanitation and PTB-criteria for hazard assessment. However, it is considered the responsibility of the industry to perform PTB-exercises: RIVM/CSR may be only involved in the methodology assessment.

5.2 Criteria for accepting and selecting QSARs

Two types of selection criteria are applied: scientific and business economical criteria.

5.2.1 Scientific criteria

Criteria to determine whether a QSAR is acceptable for use within the risk assessment process strongly depend on factors as the endpoint under consideration, the method used to generate the QSAR and the domain of the QSAR. An overview is given by Hansen et., 1997:

- **Endpoint**: Exact endpoint should be described and the experimental error has to be judged (standard deviation of measurements in the training set must be given). The experimental error should be constant over the range of the model. Units in which the endpoint results are measured should be clearly stated, and if applicable, nominal or measured concentrations should be recorded. It has to be noted that some QSARs do not have a clearly defined endpoint, e.g. biodegradation.
- **Test method**: The method used for development of the QSAR should be clearly described (e.g. details of species tested or tested soil types) and should be checked for acceptable protocol.
- **Model**: The model should reflect the underlying process described by the QSAR (e.g physico-chemical or biological interactions). The technique used to generate the model should be clearly stated (methodology, statistical package etc.) and should be appropriate. It should be checked if the model has been properly validated.

- **Descriptors**: For the derivation and application of QSARs, it is a prerequisite to realise which descriptors are mutually related, i.e. contain similar information. The assemblage of interrelated parameters in a model to satisfy statistical ambitions will not yield a meaningful relationship but will obscure its significance and predictive power. The descriptor variables used in the model should be well defined and should be reproducible. The accuracy of the determination of the descriptor variables should be available, as should the data to generate the descriptor variables. It should be stated how many (and which) variables were considered when developing the QSAR and how many (and which) variables are present in the final QSAR. The exact source of the data used for the descriptor variables should be given.
- **Domain:** The exact domain of definition of the model should be stated, which include the exact structural rules defining the group of substances for which the model is valid. In addition the ranges of the model parameters for which the model is valid should be given. If the domain is mainly defined by exclusion rules rather than inclusion rules, it could be an indication that the domain has been defined after building the model.
- **Validity**: Method and statistical design used for generating the training set should be given. The training set should consist of a sufficient number of chemicals, at least 4-5 compounds per parameter (variable) of the model. Explanations should be given for removing outliers from the training set. If outliers have been removed, then the reason for doing so should be checked for consistency. Cross validation of the model: for a sufficiently large number of compounds not used to derive the model, predictions have to be calculated and compared to the measured data. Only if there is adequate agreement the validity of the model can be assumed.
- **Accuracy:** The correlation coefficient for the model should be given, as well as overall statistics judging the overall validity and accuracy of the model. These statistics should preferably include the estimated standard deviation of the prediction errors, the statistics describing the significance of the model as a whole, and the significance of the individual variables in the model, as well as estimates (and if appropriate, the estimated standard deviation) of the model parameters.

5.2.2 Business economics

Taking into account the tasks of RIVM/CSR the involvement of CSR is in applying QSARS in hazard and risk assessment rather than in deriving them. Generally, every CSR co--worker is assumed to have a level of expertise needed for performing proper and reliable assessments that are required. However, it may be questioned whether the scientific level of QSAR approaches (3D, NN) that are in development goes with the basics. If not, investment in gaining and maintaining expertise are only worthwhile when broad application in assessments is foreseen and even then the involvement of only a few co-workers may be appropriate.

5.3 Proposal for QSAR application within RIVM/CSR

- QSARs for environmental exposure assessment may be used for all RIVM/CSR risk assessment categories. An overview of the recommended equations is given in Chapter 4.1.

- QSARs for environmental effects assessment may be useful in checking experimental data in all types of risk assessments, but are considered most useful in the framework of deriving INS- and I-values in case:
  - experimental data are found to be doubtful.
  - too few (one or two LC50 or one NOEC) or no experimental data are available.

It is concluded that [a] the most promising QSARs are still in a developmental stage and [b] QSARs for effect assessment based on other descriptors than logKow, require a high level of expertise in this field to ensure proper application. Taking into account the supply/demand ratio as well, it is recommended to limit QSAR application in such a way that the method is easy to perform and the results are reproducable.
The various aspects of effects assessment based on QSARs have been subject to discussion with experts (see Appendix II), leading to the following procedure recommended for toxicity estimation within RIVM/CSR and is presented in flow diagram 5.3.1:

- Use the ECOSAR program as a basis for environmental effect assessments. ECOSAR is easy to use and requires only a CAS-number or a SMILES notation as input parameter. ECOSAR generates automatically the chemical class to which the chemical compound belongs. This ECOSAR classification makes it easy to classify the chemical concerned in one of the 4 chemical classes proposed by Verhaar et al. (1992). In Appendix II a scheme is given how the different chemical classes proposed in ECOSAR can be integrated in the classification scheme of Verhaar et al. (1992).

- If no QSAR estimates are available in ECOSAR, expert judgement is required for applying other published QSARs (an overview of available QSARS is given in Chapter 4.2.2); if QSARS are available in ECOSAR certain QSARs should not be used since they are considered to be unreliable (based on less than 5 experimental data and QSAR estimates with an r^2 beneath 0.7). See Appendix II for these exclusions, and Appendix III for an example of an ECOSAR estimation record.

- In case more than one QSAR estimate is given for different species, the lowest value has to be chosen, although in some cases there may be good arguments to make an other decision.

- Because ECOSAR classifies a chemical on the basis of functional groups, it may occur that a chemical is classified as belonging to more than one chemical class. In that case it is difficult to make the choice for the most appropriate QSAR estimate. Therefore it is decided to use the lowest QSAR estimate.

- If a QSAR estimate for baseline toxicity is found to be lower than the reported experimental value, use the lower QSAR value, because a chemical is always as toxic as its baseline toxicity.

- Apply a factor 10 for extrapolation of acute toxicity QSAR values to QSAR chronic values.

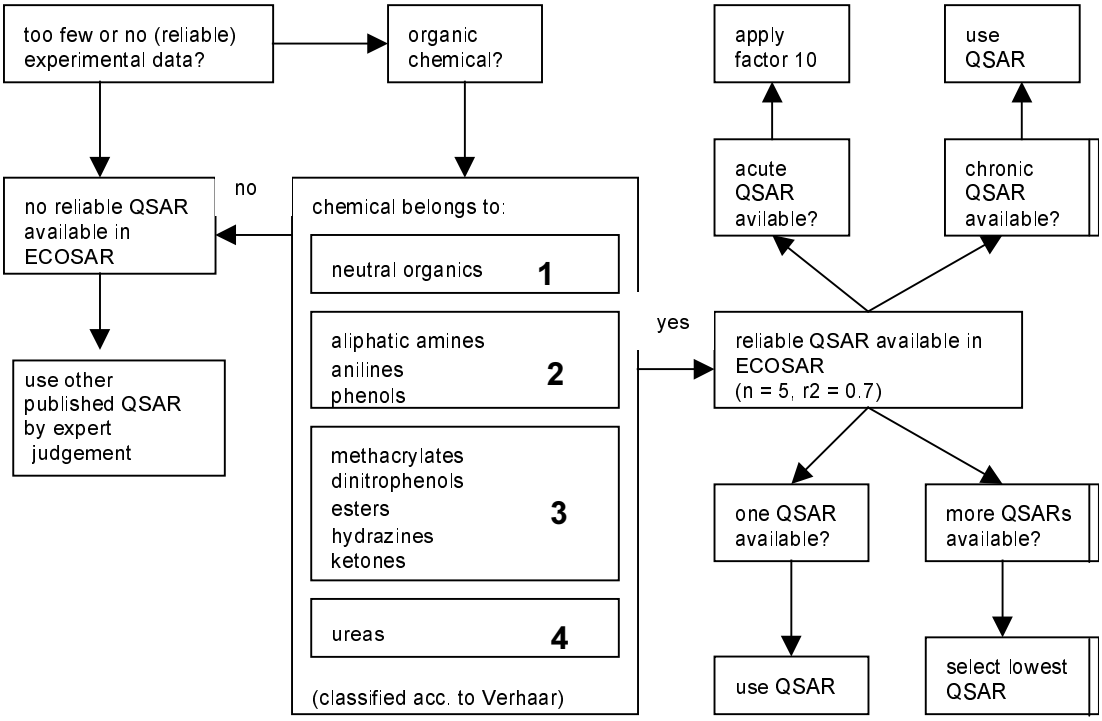


Figure 5.1: Flow-diagram for QSAR application procedure

References

Abernethy, S., A.M. Bobra, W.Y. Shiu, P.G. Wells, and D. Mackay (1986) Acute lethal toxicity of hydrocarbons and chlorinated hydrocarbons to two planktonic crustaceans: the key role of organism-water partitioning. Aquatic Toxicol., 8, 163-174

Altschuh, J., R. Bruggemann, and W. Karcher (1993) Attempts to classify QSARs with respect to their validity: vapour pressure estimation as an example. Sci. Total Environ., Suppl. 1993, 1409-1419

AQUIRE, Aquatic Infromation Retrievel, sponsored by US-EPA

ASTER (1994) Assessment Tools for the Evaluation of Risk, Developed at US-EPA, ERL-Duluth

Atkinson, R. (1987) Structure-activity relationship for the estimations of rate constants for the gas-phase reactions of hydroxyl radicals with organic compounds. Int. J. Chem. Kinet., 19, 799-828

Atkinson, R. (1988) Estimation of gas-phase hydroxyl radical rate constants for organic chemicals. Environ. Toxicol. Chem., 7, 435-442

Bearden, A.P., and T.W. Schultz (1998) Comparison of *Tetrahymena* and *Pimephales* toxicity based on mechanism of action. SAR QSAR Environ. Res., 9, 3-4, 127-153

Barron, M.G., M.J. Anderson, J. Lipton, and D.G. Dixon (1997) Evaluation of critical body residue QSARs for predicting organic chemical toxicity to aquatic organisms. SAR QSAR Environ. Res., 6, 47-62

Basak, S.C., G. D. Grunwald, G.E. Host, G.J. Niemi, and S.P. Bradbury (1998) A comparative study of molecular similarity, statistical, and neural methods for predicting toxic modes of action. Environ. Toxicol. Chem., 17, 6, 1056-1064

Benoit-Guyod, J.L., C. Andre, G. Taillandier, J. Rochat, and A. Boucherle (1984) Toxicity and QSAR of chlorophennols on *Lebistes reticulatus*. Ecotoxicol. Environ. Saf., 8, 227-235

Bintein, S., J. Devillers, and W. Karchner (1993) Nonlinear dependence of fish bioconcentration on n-octanol/water partition coefficient. SAR QSAR Environ. Res., 1, 29-39

BKH/Haskoning (1998) Selection of PTBs phase 3. Ministry of Housing, Physical Planning and the Environment

Bobra, A.M., W.Y. Shiu, and D. Mackay (1983) Acute toxicity of fresh and weathered crude oils to *Daphnia magna*. Chemosphere, 12, 1137-1149

Bobra, A.M., W.Y. Shiu, and D. Mackay (1984) Structure-activity relationships for the toxicity of hydrocarbons, chlorinated hydrocarbons and oils to *Daphnia magna*. In: QSAR in Environmental Toxicology. Kaiser, K.L.E., Ed. Kluwer, Dardrecht, pp. 3-16

Boyd, E.M., A.A. Meharg, J. Wright, and K. Killham (1998) Toxicity of chlorobenzenes to a lux-marked terrestrial bacterium, *Pseudomonas fluorescens*. Environ. Toxicol. Chem., 17, 11, 2134-2140.

Bradbury, S.P. (1994) Predicting modes of toxic action from chemical structure: An overview. SAR and QSAR in environmental research, 2, 89-104*

Briens, F., R. Bureau, S. Rault, and M. Robra (1995) Applicability of CoMFA in ecotoxicology: A critical study on chlorophenols. Ecotoxicol. Environ. Saf., 31, 37-48

Briens, F., R. Bureau, and S. Rault (1999) Applicability of CATALYST in ecotoxicology, a new promising tool for 3D-QSAR: Study of chlorophenols. Ecotoxicol. Environ. Saf., 43, 241-251

Brooke, L.T., D.J. Call, D.L. Geiger, S.H. Porier, and C.E. Northcott (1984, 1985, 1986, 1988) Acute toxicities of organic chemicals to fathead minnow (*Pimephales promelas*). Centre for Lake Superior Environmental Studies, University of Wisconsin-Superior, volume 1-4

Bureau, R., J.C. Faucon, J. Faisant, F. Briens, and S. Rault (1997) Applicability of the free energies of solvation for the prediction of ecotoxicity: Study of chlorophenols. SAR QSAR Environ. Res., 6, 163-181

Cajina-Quezada, M., and T.W. Schultz (1990) Structure toxicity relationships for selected weak acid respiratory uncouplers. Aquatic Toxicol., 17, 239-252

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

Canton, J.H., and W. Slooff (1979) A proposal to classify compounds and to establish water quality criteria based on laboratory data. Ecotox. Environ. Saf. 3, 126-132

Chiou, C.T., V.H. Freed, D.W. Schmedding, and R.L. Kohnert (1977) Partition coefficients and bioaccumulation of selected organic chemicals. Environ. Sci. Technol.,11, 475-478

Clements, R.G., J.V. Nabholz, and M. Zeeman (1996) ECOSAR: A computer program for estimating the toxicity of industrial chemicals based on structure-activity relationships. Environmental Effects Branch, Health and Environmental Review Division, US EPA, Washington, DC

Clements, R.G., J.V. Nabholz, M.G. Zeeman, and C.M. Auer (1995)
The application of structure-activity relationships (SARs) in the aquatic toxicity evaluation of discrete organic chemicals. SAR QSAR Environ. Res., 3, 203-215

Connell, D.W., and D.W. Hawker (1988) Use of polynominal expressions to describe the bioconcentration of hydrophobic chemicals in fish. Ecotoxicol. Environ. Saf., 16, 242-257

Connell, D.W., and R.D. Markwell (1990) Bioaccumulation in the soil earthworm system. Chemosphere, 20, 91-100

Crommentuijn, G.H., E.J. van de Plassche, and J.H. Canton (1994) Guidance document on the drivation of ecotoxicological criteria for serious soil contamination in view of the Intervention Value for soil and groundwater. National Institute of Public Health and the Environment, Bilthoven, The Netherlands. Report no. 950011003

Cronin, M.T.D., and T.W. Schultz (1996) Structure-toxicity relationships for phenols to *Tetrahymena pyriformis*. Chemosphere, 32, 8, 1453-1468

Cronin, M.T.D., and T.W. Schultz (1997) Validation of *Vibrio fisheri* acute toxicity data: mechanism of action-based QSARs for non-polar narcotics and polar narcotic phenols. Sci. Total Environ. 201, 75-88

Cronin, M.T.D., and T.W. Schultz (1998) Structure-activity relationships for three mechanisms of action of toxicity to *Vibrio fisheri*. Ecotoxicol. Environ. Saf., 39, 65-69

Cronin, M.T.D., Y.H. Zhao, and R.L. Yu (2000) pH-Dependence and QSAR analysis of the toxicity of phenols and anilines to *Daphnia magna*. Environ. Toxicol., 15, 140-148

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

Deneer, J.W., W. Seinen, and J.L. Hermens (1988a) The acute toxicity of aldehydes to the guppy. Aquatic Toxicol., 12, 185-192

Deneer, J.W., T.L. Sinnige, W. Seinen, and J.L.M. Hermens (1988b) A quantitative structure-activity relationship for the acute toxicity of some epoxy compounds to the guppy. Aquatic toxicol., 13, 195-204

Devillers, J., D. Domine, and W. Karcher (1995) Estimating n-octanol water partition coefficients from the autocorrelation method. SAR QSAR Environ. Res., 3, 301-306

Devillers, J., and P. Chambon (1986) Acute toxicity and QSAR of chlorophenols on *Daphnia magna*. Bull. Environ. Contam. Toxicol., 37, 599-605

Devillers, J., and D. Domine (1999) A noncongenric model for predicting toxicity of organic molecules to *Vibrio fisheri*, SAR QSAR Environ. Res., 10, 61-70

De Wolf, W. J.H. Canton, J.W. Deneer, R.C.C. Wegman, and J.L.M. Hermens (1988) Quantitative structure-activity relationships and mixture-toxicity of alcohols and chlorohydrocarbons: reproducibility of effects on growth and reproduction of *Daphnia magna,* Aquatic Toxicol., 12, 39-49

De Zwart, D (2000) Observed regularities in SSDs for aquatic species (in press)

ECETOC, Technical Report No.74: QSARs in the assessment of the environmental fate and effects of chemicals, June 1998

Eldred, D.V., C.L. Weikel, P.C. Jurs, and K.L.E. Kaiser (1999) Prediction of fathead minnow acute toxicity of organic compounds from molecular structure. Chem. Res. Toxicol., 12, 7, 670-678

Environmental categorisation for persistence, bioaccumulation and inherent toxicity of substances on the domestic substances list using QSARs. Results of an international QSAR workshop, final reportt, July 2000

EPIWIN (1997) Estimation Program Interface for Microsoft Windows 3.1, Syracuse Research Corp. North Syracuse, New Yersey

Eriksson, L., H.J.M. Verhaar, and J.L.M. Hermens (1994) Multivariate characterisation and modelling of the chemical reactivity of epoxides. Environ. Toxicol. Chem., 13, 5, 683-691

EU-DG-XII Project: QSAR for predicting fate and effects of chemicals in the environment. Contract # EV5V-CT92-0211, July 12, 1995

EUDRALEX, Guidelines veterinary medicinal products. European commission, 1999

Feijtel, T.C.J. (1995) Evaluation of the use of QSARs for priority setting and risk assessment. SAR and QSAR in environmental research, 3, 237-245*

Freidig, A.P., H.J.M. Verhaar, and J.L.M. Hermens (1999) Quantitative structure-property relationships for the chemical reactivity of acrylates end methacrylates, Environ. Toxicol. Chem., 18, 6, 1133-1139

Girones, X., L. Amat, and R. Carbo-Dorca (1999) Using Molecular Quantum Similarity measures as descriptors in Quantitative Structure-Toxicity Relationships. SAr QSAR Environ. Res., 10, 545-556

Gombar, V.K. (1987) Quantitative structure-activity relationships studies: acute toxicity of environmental contaminants. In: QSAR in Environmental Toxicology. Kaiser, K.L.E., Ed. Kluwer, Dordrecht, pp. 125-133

Govers, H., C. Ruepert, and H. Aiking (1984) Quantitative structure-activity relationships for polycyclic aromatic hydrocarbons: correlation between molecular connectivity, physicochemical properties, bioconcentration and toxicity in *Daphnia pulex*. Chemosphere, 13, 227-236

Grain, C.F. (1990) Vapour pressure, in Handbook of chemical property estimation methods (eds. W.J. Lyman, W.F. Reehl, and D.H. Rosenblatt) American Chemical Society, Washington, DC

Güsten, H., L. Klasinc, and D.Maric (1984) Prediction of the abiotic degradability of organic compounds in the troposphere. J. Atmos. Chem., 2, 83-93

Hall, L.H., and L.B. Kier (1984) Molecular connectivity of phenols and their toxicity to fish. Bull. Environ. Contam. Toxicol. 32, 354-362

Hall, L.M., E.L. Maynard, and L.B. Kier (1989) Structure-activity relationship studies on the toxicity of benzene derivatives: III Predictions and extension to new substituents. Environ. Toxicol. Chem., 8, 431-436

Hansch, C., J.E. Quinlan, and G.L. Lawrence (1968) The linear-free energy relationship between partition coefficients and the aqueous solubility of organic liquids. J. Org. Chem., 33, 347-350

Hansch, C., and W.J. Dunn III (1972) Linear relationships between lipophilic character and biological activity in drugs. J. Pharmacol. Sci., 61, 1-19

Hansch, C., and A.J. Leo (1979) Substituent constants for correlation analysis in chemistry and biology. Wiley, New York

Hansen, B., H. Loonen, and W. Karchner (1997) Use of QSARs for risk assessment of chemicals in the European Union. In: Quant. Sruct.-Act. Relat. Environ. Sci.-VII, Proc. QSAR 96; SETAC Press: Pensacola, Fla, 413-424. Editor: Chen, F; Schueuermann, G.

Hansen, B.G., A. van Haelst, K. van Leeuwen, and P. van der Zandt (1999) Priority setting for existing chemicals: European Union Risk ranking method. Environ. Toxicol. Chem., 18, 4, 772-779

Hermens, J.L.M., H. Canton, P. Janssen, and R. De Jong (1984a) Quantitative structure-activity relationships and toxicity studies of mixtures of chemicals with anaestetic potency: Acute lethal and sublethal toxicity to *Daphnia magna*. Aquatic Toxicol., 5, 143-154

Hermens, J.L.M., P. Leeuwangh, and A. Musch (1984b) Quantitative structure-activity relationships and mixture toxicity studies of chloro-and alkylanilines in an acute lethal toxicity level to the guppy *(Poecilia reticulata*) Ecotoxicol. Environ. Saf., 8, 388-394

Hermens, J., S. Balaz, J. Damborsky, W. Karchner, M. Müller, W. Peijnenburg, A. Sabljic, and M. Sjöström (1995) Assessment of QSARs for predicting fate and effects of chemicals in the environment: An international European project. SAR and QSAR in environmental research, 3, 223-236.

Hiromatsu, K., Y. Yakabe, K. Katagiri, and T. Nishihara (2000) Predicting for biodegradability of chemicals by an empirical flowchart. Chemosphere, 41, 1749-1754

Ikemoto, Y., K. Motoba, t. Suzuki, and M. Uchida (1992) Quantitative structure-activity relationships of nonspecific and specific toxicants on several organism species. Environ. Toxicol. Chem., 11, 931-939

Isnard, P., and S. Lambert (1989) Aqueous solubility and n-octanol/water partition coefficient correlations. Chemosphere, 18,1837-1853

IUCT (1992) Handbuch zum SAR-Program Version 3.0, Fraunhofer Institute für Umweltchemie und Ökotoxikologie, Schmallenberg.

Jager, D.T., and T. Hamers (1997) Estimation methods for bioaccumulation in risk assessment of organic chemicals. National Institute of Public Health and the Environment, Bilthoven, The Netherlands, Report no. 679102 013

Jager, D.T. (1998) Mechanistic approach for estimating bioconcentration of organic chemicals in earthworms (oligochaeta). Environ. Toxicol. Chem. 17, 10, 2080-2090

Jin, L.., J. Dai, P. Guo, L. Wang, and Z. Wei (1998) Quantitative structure-activity relationships for benzaldehydes to *Daphnia magna*. Chemosphere, 37, 1, 79-85

Kaiser, K.L.E., and S.P. Niculescu (1999) Using Probabilistic Neural Networks to model the toxicity of chemicals to the fathead minnow (*Pimephales promelas*) : A study based on 865 compounds. Chemosphere, 38, 14, 3237-3245

Kalf, D.F., B.J.W.G. Mensink, M.H.M.M. Montforts (1999) Protocol for derivation of harmonised Maximum Permissible Concentrations (MPCs) National Institute of Public Health and the Environment, Bilthoven, The Netherlands, report no. 601506001

Karchner, W., B.G. Hansen, C. van Leeuwen, P. Wagner, and C. Auer (1995) Predictions for existing chemicals – A multilateral QSAR project. SAR and QSAR in environmental research, 3, 217-221*

Karchner, W., and S. Karabunarliev (1996) The use of computer based structure-activity relationships in the risk assessment of industrial chemicals. J. Chem. Inf. Comput. Sci., 36, 672-677

Kenaga, E.E., and C.A. Goring (1980) Relationship between water solubility, soil sorption, octanol-water partitioning and bioconcentration of chemicals in biota, in Aquatic Toxicology, (eds. J.G. Eaton, P.R. Parrish, and A.C. Hendricks) STP707 ASTM, Philadelphia, PA.

Klamt, A. (1993) Estimation of gas-phase hydroxyl radical constants of organic compounds from molecular orbital calculations. Chemosphere, 26, 1273-1289

Klopman, G., R. Saiakhov, H.S. Rosenkranz, and J.L.M. Hermens (1999) Multiple-Computer-Automated Structure Evaluation program study of aquatic toxicity 1: Guppy. Environ. Toxicol. Chem., 18,11, 2497-2505

Könemann, W.H., and C. van Leeuwen (1980)
Toxicokinetics in fish: accumulation and elimination of six chlorobenzenes in guppies. Chemosphere, 9, 3-19

Könemann, W.H. (1981a) Quantitative structure-activity relationships in fish toxicity studies. Part I. Relationship for 50 industrial pollutants, Toxicology, 19, 209-221

Könemann, H. (1981b) Fish toxicity tests with mixtures of more than two chemicals: a proposal for a quantitative approach and experimental results. Toxicology, 19, 229-238

Könemann H., and A. Musch (1981c) Quantitative structure-activity relationships in fish toxicity studies. Part 2: The influence of pH on the QSAR of chlorophenols. Toxicology, 19, 223-228

Landrum, P.F., S.W. Fisher, H. Hwang, and J. Hickey (1999) Hazard evaluation of ten organophosphorus insecticides against the midge, *Chironomus riparius* via QSAR. SAR QSAR Environ. Res., 10, 423-450

Lange A.W., and K. Vormann (1995) Experiences with the application of QSAR in the routine of the notification procedure. SAR QSAR Environ. Res., 3, 171-177*

Langenberg, J.H., W.J.G.M. Peijnenburg. and E. Rorije (1996) On the usefulness and reliability of existing QSBRs for risk assessment and priority setting. SAR QSAR Environ. Res., 5, 1-16

Lipnick, R.L., K.R. Watson, and A.K.S strausz (1987) A QSAR study of the acute toxicity of some industrial organic chemicals to goldfish. Narcosis, electrophile and proelectrophile mechanisms. Xenobiotica, 17, 1011-1025

Lipnick, R.L. (1995) Computational chemistry in environmental toxicology QSAR. SAR and QSAR in environmental research, 4, 125-130

Lu, X., S. Tao. H. Hu, and R.W. Dawson (2000) Estimation of bioconcentration factors of nonionic organic compound in fish by molecular connectivity indices and polarity correction factors. Chemosphere, 41, 1675-1688

Luttik, R., and J.H.M. de Bruijn (1998) Risicobeoordeling van stoffen voor mens en milieu—Een anlyse van hiaten in de beoordelingsmethodiek. National Institute of Public Health and the Environment, Bilthoven, The Netherlands, CSR adviesrapport nr 05931A00

Lyman (1990) Handbook of chemical property estimation methods (eds. W.J. Lyman, W.F. Reehl, and D.H. Rosenblatt) American Chemical Society, Washington, D.C.

Nabholz, J.V., and R.D. Platz (1990) Generic environmental hazard assessment- Aliphatic amines. Environmental Effects Branch, Health and Environmental Review Division (TS-796), Office of Toxic Substances, US EPA, Washington, DC

Nenzda, M., and W. Klein (1990) Comperative QSAR study on freshwater and estuarine toxicity. Aquatic Toxicol., 17, 63-74

Nenzda, M. (1991) QSARs of bioconcentration: validity assessment of logPow/logBCF correlations, in Bioaccumulation in aquatic systems (eds. R. Nagel, r. Loskill), VCH, Weinheim

Nenzda, M., and C.L. Russom (1991) QSAR modelling of the ERL-D fathead minnow acute toxicity database. Xenobiotica, 21, 147-170

Nenzda, M. (1998) Structure-Activity Relationships in environmental sciences. Chapman & Hall Ecotoxicology Series 6.

Newsome, L.D., D.E. Johnson, R.L. Lipnick, S.J. Broderius, and C.L. Russom (1991) A. QSAR study of the toxicity of amines to the fathead minnows. Sci. Total. Environ. 109/110, 537-551

Mackay, D. (1982) Correlation of bioconcentration factors. Environ. Sci. Technol., 16, 274-276

Mackay, D., A. Bobra, D.W. Chan, and W.Y. Shiu (1982) Vapour pressure correlations for low volatility environmental chemicals. Environ. Sci. Technol., 16, 645-649

Mishra, D.S., and S.H. Yalkowsky (1991) Estimation of vapour pressure of some organic compounds. Ind. Eng. Chem. Res., 30, 1609-1612

Müller, M., and W. Klein (1992) Comperative evaluation of methods predicting water solubility for organic compounds. Chemosphere, 25, 769-782

Rekker, R.F. (1977) The hydrophobic fragmental constant, Elsevier Scientific, New York, NY

Ribo, J.M., and J.L.E. Kaiser (1983) Effects of selected chemicals to photoluminescent bacteria and their correlations with acute and sublethal effects on other organisms. Chemosphere, 12, 1421-1442

Riederer, M. (1990) Estimating partitioning and transport of organic chemicals in the foliage/atmosphere system: Discussion of a fugacity-based model. Environ. Sci. Technol., 24, 829-837

Robert, D., and R. Carbo-Dorca (1999) Aromatic compounds aquatic toxicity QSAR using molecular quantum similarity measures. SAR QSAR Environ. Res., 10, 401-422

Roberts, D.W. (1987) An analysis of published data on fish toxicity of nitrobenzene and aniline derivatives. In: QSAR in Environmental Toxicology. Kaiser, K.L.E., Ed. Kluwer, Dordrecht, pp. 295-308

Roghair, C.J., A. Buijze, M.P.A. Huys, M.A.H. Wolters-Balk, E.S.E. Yedema, and J.L.M. Hermens (1996)Toxicity and toxicokinetics for benthic organisms;II: QSARs for base-line toxicity to the midge *Chironomus riparius* and the tubificid oligachaete worm Branchiura *sowerby*. National Institute of public Health and the Environment, Bilthoven, The Netherlands and Research Institute Toxicology, Utrecht, The Netherlands. Report no. 719101 026.

Rorije, E., M. Müller, and W.J.G.M. Peijnenburg (1997) Prediction of environmental degradation rates fo High Production Volume Chemicals (HPVC) using Quantitative Structure-Activity Relationships. National Institute of public Health and the Environment, Bilthoven, The Netherlands. Report no. 719101 030.

Rorije, E., H. Loonen. M. Müller, G. Klopman, and W.J.G.M. Peijnenburg (1999) Evaluation and application of models for the prediction of ready biodegradability in the MITI-test. Chemosphere, 38, 6, 1409-1417

Russom, C.L. S. P. Bradbury, S.J. Broderius (1997) Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). Environ. Toxicol. Chem., 16, 5, 948-967

Saarikoski, J., and M. Viluksela (1982) Relation between physicochemical properties of phenols and their toxicity and accumulation in fish. Ecotoxicol. Environ. Saf., 6, 501-512

Sabljic, A (1983) Quantitative structure-activity relationships of chlorinated compounds: a molecular connectivity investigation, Bull. Environ.Contam. Toxicol., 30, 80-83

Sabljic, A., and H. Güsten (1990) Predicting the night-time NO3 radical reactivity in the troposphere. Atmosph. Environ., 24A, 73-78

Schultz, T.W., G. W. Holcombe, and G.L. Phipps (1986) Relationships of Quantitative structure-activity to comparative toxicity of selected phenols on the *Pimephales promelas* and *Tetrhymena pyriformis* test system. Ecotoxicol. Environ. Saf., 12, 146-153

Schultz, T.W., T.S. Wilke, S.E. Bryant, and L.M. Hosein (1991) QSARs for selected aliphatic and aromatic amines. Sci. Total Environ., 109/110, 581-587

Schultz, T.W., and M.T.D. Cronin (1997) Quantitative structure-activity relationships for weak and respiratory uncouplers to *Vibrio fisheri* .Environ. Toxicol. Chem. 16, 2, 357-360

Schultz, T.W., G.D. Sinks, and M.T.D. Cronin (1997) Identification of toxic action of phenols to *Tetrahymena pyriformis* from molecular descriptors. Quant. Sruct.-Act. Relat. Environ. Sci.-VII, Proc. QSAR 96; SETAC Press: Pensacola, Fla, 329-342. Editor: Chen, F; Schüürmann, G.

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

Seward, J.R., and T.W. Schultz (1999) QSAR analysis of the toxicity of aliphatic carboxylic acids and salts to *Tetrahymena pyriformis*. SAR QSAR Environ. Res., 10, 557-567

Sijm, D., E. Hulzebos, and W. Peijnenburg (1999) Estimating the PTB-profile. National Institute of public Health and the Environment, Bilthoven, The Netherlands. Report no. 601503 016

Sijm, D., and R.Luttik (1999) Hiaten in de risicobeoordeling: contactpersonen buiten het RIVM. CSR adviesrapport 006836A00

Sixt, S., and J. Altschuh (1997) Prediction of luminescent bacteria toxicity using quantum chemical descriptors: test of a classification scheme. Quant. Sruct.-Act. Relat. Environ. Sci.-VII, Proc. QSAR 96; SETAC Press: Pensacola, Fla, 343-362. Editor: Chen, F; Schueuermann, G.

Slooff, W. (1992) Ecotoxicological effect assessment: Deriving Maximum Tolerable Concentrations (MTC) from single-species toxicity data. National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands, report no. 71910218

Solbé, J., U.Mark, B. Buyle, W. Guhl, T. Hutchinson, P. Kloepper-Sams, R. Lange, R. Munk, N. Scholtz, W. Bontonck, and H. Niessen (1998) Analysis of the ECETOC aquatic toxicity (EAT) database I – General introduction. Chemosphere, 36, 1, 99-113

Trapp, S., and M. Matthies (1995) Generic one-compartment model for the uptake of organic chemicals by foliar vegetation. Environ. Sci. Technol., 29, 2333-2338

TGD, Technical Guidance Documents in support of: The comission Directive 93/67/EEC on risk assessment for new notified substances and The Commission Regulation (EC) 1488/94 on risk assessment for existing substances. ECB, Ispra, Italy, 19th April, 1996.

Todeschini, R., and P. Gramatica (1997) The WHIM theory: New 3D molecular descriptors for QSAR in environmental modelling. SAR QSAR Environ. Res., 7, 89-115

Tuppurainen , K., and J. Ruuskanen (2000) Electronic eigenvalue (EEVA): a new QSAR/QSPR descriptor for electronic substituent effects based on molecular orbital energies. A QSAR approach to the Ah receptor binding affinity of polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs). Chemosphere, 41, 843-848

Urrestarazu Ramos, E., W.H.J. Vaes, P. Mayer, and J.L.M. Hermens (1999) Algal growth inhibition of *Chlorella pyrenoidosa* by polar narcotic pollutants: toxic cell concentrations and QSAR modelling

Urrestarazu Ramos, E., W.H.J. Vaes, H.J.M. Verhaar, and J.L.M. Hermens (1998) Quantitative structure-activity relationships for the aquatic toxicity of polar and nonpolar narcotic chemicals. J. Chem. Inf. Comput. Sci., 38, 845-852

Urrestarazu Ramos, E., W.H.J. Vaes, H.J.M. Verhaar, J.L.M. Hermens (1997) Polar narcosis: Designing a suitable training set for QSAR studies. Environ. Sci. & Pollut. Res., 4, 2, 83-90

Vaes, W.H.J., E. Urrestarazu Ramos, H.J.M. Verhaar, and J.L.M. Hermens (1998) Acute toxicity of nonpolar versus polar narcosis: Is there a difference? Environ. Toxicol. Chem., 17, 7, 1380-1484

Valvani, S.C., S.H. Yalkowsky, and T.J. Rosenman (1981) Solubility and partitioning IV: aqueous solubility and octanol-water partition coefficients of liquid non-electrolytes. J. Pharmacol. Sci., 70, 502-507

Van Gestel, C.A.M., and W.C. Ma (1988) Toxicity and bioaccumulation of chlorophenols in earthworms in relation to bioavailability in soil. Ecotoxicol. Environ. Saf., 15, 289-297

Van Leeuwen, C.J. D.M. Adema, and J. Hermens (1990) Quantitative structure activity relationships for fish early fife stage toxicity. Aquatic Toxicol., 16, 321-334

Van Leeuwen, C.J., P.T.J. van der Zandt, T. Aldenberg, H.J.M. Verhaar, and J.L.M. Hermens (1992) Application of QSARs, extrapolation and equilibrium partitioning in aquatic effects assessment. I. Narcotic industrial chemicals. Environ. Toxicol. Chem., 11, 267-282

Veith, G.D., D.L. Defoe, and B.V. Bergsted (1979) Measuring and estimating the bioconcentration factor of chemicals in fish. J. Fish. Res. Board. Can., 36, 1040-1048

Veith, G.D., K.J. Macek, S.R. Petrocelli, and J. Carrol (1980) An evaluation of using partititon coefficients and water solubility to estimate bioconcentration factors for organic chemicals in fish, in Aquatic Toxicology (eds, J.G. Eaton, P.R. Parrish, and A.C. Hendricks). STP 707, ASTM, Philadelphia, PA

Veith, G.D., and S.J. Broderius (1987) Structure-toxicity relationships for industrial chemicals causing type II narcosis syndrome. In: QSAR in Environmental Toxicology. Kaiser, K.L.E., Ed. Kluwer, Dordrecht, pp. 385-391

Veith, G.D. and O.G. Mekenyan (1993) A QSAR approach for estimating the aquatic toxicity of soft electrophiles. Qant. Struct-Act. Relat., 12, 349-356

Verhaar , H.J.M., C.J. van Leeuwen, and J.L. M. Hermens (1992) Classifying environmental pollutants. 1: Activity relationships for prediction of aquatic toxicity. Chemosphere, 25, 4, 471-491.

Verhaar, H.J.M., C.J. Van Leeuwen, J. Bol, and J.L.M. Hermens (1994) Application of QSARs in risk management of existing chemicals. SAR QSAR Environ. Res., 2, 39-58

Verhaar, H.J.M., E. Urrestarazu Ramos, and J.L.M. Hermens (1996) Classifying environmental pollutants. 2: Separation of Class 1 (baseline toxicity) and class 2 ('polar narcosis") type compounds based on chemical descriptors. J. Chemometr., 10, 149-162

Verhaar, H.J.M., E. Rorije, H. Borkent, W. Seinen, and J.L.M. Hermens (1996) Modelling the nucleophilic reactivity of small organochlorine electrophiles: A mechanistically based quantitative structure-activity relationship. Environ. Toxicol. Chem., 15, 6, 1011-1018

Verhaar, H.J.M., J. Solbé, J. Speksnijder, C.J. van Leeuwen, and J.L.M. Hermens (2000) Classifying environmental pollutants: Part 3. External validation of the classification system. Chemosphere, 40, 875-883

Versteeg, D.J., D.T. Stanton. M.A. Pence, and C. Cowan (1997) Effects of surfactants on the rotifer *Brachionus calyciflorus*, in a chronic toxicity test and in the developement of QSARS. Environ. Toxicol. Chem., 15, 5, 1051-1058

Vighi, M., and D. Calamari (1987) A triparametric equation to describe QSARs for heterogenous chemical substances. Chemosphere, 16, 1043-1051

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

Vighi, M. (1998) The use of QSARs for heterogeneous chemicals substances: meaning, predictive capability, and practical applications. Biotherapy 11, 97-104

Wang, X., Y. Dong, S.Xu, L. Wang, and S. Han (2000) Quantitative structure-activity relationships for the toxicity to the tadpole *Rana japonica* of selected phenols. Bull. Environ. Contam. Toxicol., 64, 859-865

Warne, M.A., E.M. Boyd, A.A. Meharg, D. Osborn, K,Killham, J.C. Lindon, and J.K. Nicholson (1999) Quantitative structure-toxicity relationships for halobenzenes in two species of bioluminescent bacteria, *Pseudomonas fluorescens* and *Vibrio fisheri*, using an atom-centered semi-empirical molecular-orbital based model. SAR QSAR Environ. Res., 10, 1, 17-38

Weyers, A., B. Sokull-Klüttgen, J. Baraibar-Fentanes, and G.Vollmer (2000) Acute toxicity data: A comprehensive comparison of results of fish, *Daphnia*, and algae tests with new substances notified in the European Union. Enviton. Toxicol. Chem., 19, 7, 1931-1933

Wong, P.T.S., Y.K. Chau, J.S. Rhamey, and M. Docker (1984). Relationship between water solubility of chlorobenzenes and their effects on a freshwater alga. Chemosphere, 9, 991-996

Wong, D.C.L., P.B. Dorn, and E.Y. Chai (1997) Acute toxicity and structure-activity relationships of nine alcohol ethoxylate surfactants to fathead minnow and *Daphnia magna*.
Environ. Toxicol. Chem., 16, 9, 1970-1976

Xu, S., and N. Nirmalakhandan (1998) Use of QSAR models in predicting joint effects in multi-component mixtures of organic chemicals. Wat. Res., 32, 8, 2391-2399

Zeeman, M., C.M. Auer, R.G. Clements, J.V. Nabholtz, and R.S. Boethling (1995) US EPA regulatory perspectives on the use of QSAR for new and existing chemical evaluations. SAR QSAR Environ. Res., 3, 179-201

APPENDIX I: Classification of organic chemicals according to Verhaar et al. (1992)

Only those organic compounds that consist of C, H, N, O, S and/or halogens (not iodine) are considered for inclusion in classes 1 to 3.

To classify as belonging to class 1, 2 or 3, chemicals should:

$0 < \log Kow = 6$ [1] AND $MW < 600$ [2]

Class 1 type compounds (narcosis or baseline toxicity)

To classify as class 1 type compound, chemicals should:

NOT contain iodine [3] or ionic groups,

AND

contain only C & H

OR

IF containing only C, H, & halogen

be acyclic NOT containing halogen at β-positions from unsaturations (e.g. allylic/propargylic halogens) **(A)**

OR

be monocyclic compounds substituted with halogens
be monocyclic compounds that are unsubstituted or substituted with acyclic structures containing only C & H, or complying with rule A[4]
be polycyclic compounds that are unsubstituted ot substituted with acyclic structures containing only C & H, or complying with rule A[5]

OR

IF containing C, H & O

be linear ethers or monocyclic mono-ethers, but NOT epoxides or peroxides
be aliphatic alcohols, but NOT allylic/propargylic alcohols
be alcohols with aromatic moieties, bu NOT phenols or benzylic alcohols
be ketones, but NOT α,β-unsaturated ketones (e.g. 1-butenone or acetophenone)

OR

IF containing C, H & N

be aliphatic secondary or tertiary amines

OR

IF containing C, H, O & halogen
are halogenated compounds that comply with rule 1.5, but NOT α or β halogen-substituted compounds

N.B.: It may be possible that some compounds, that are known to be more toxic than baseline toxicity, do classify as class 1 compounds [6] . If this be the case, DO NOT treat these compounds as baseline toxicants.

---

[1]Compounds with a logKow lower than 0 will not be considered, due to the unrealistically high effect concentrations that will be predicted by using narcosis type QSARs. For compounds acting through a nonpolar narcosis type mode of action, it is considered unlikely that they would exhibit acute toxic action towards biota in aqueous environments because of this. Compounds that have a logKow higher than 6 do not normally exhibit acute toxicity. The most plausible explanation for this observation is that, since logKow is considered to be a parameter describing the kinetics of uptake of chemicals from water, chemicals that have logKow's in this range are generally taken up too slowly to show acute toxic effects. Furthermore some of these high logKow compounds are simply too bulky to be taken up through membranes. An example of this is tetradecanol, which can be considered to be essentially nontoxic in experiments that measure acute toxicity. Of course this rule does not mean that compounds
with logKow values that lie outside this range are to be considered nontoxic, but only that we do not recommend modeling their toxicity using narcosis-type QSAR equation.

[2]Generally speaking, compounds having MW of over 600 Daltons are too bulky to be taken up across membranes. Because of the fact that a high proportion of all chemicals with a MW higher than 600 Daltons are compounds acting by specific mechanism anyway, we define an ad hoc limit of applicability of narcosis type QSAR equations at MW = 600 Daltons.

[3]Organic compounds containing covalently bound F are to be considered equivalent with H-analogues; but please note that F-compounds are non-metabolizable if F substitutes for metabolically important H-atoms. This can give rise to chronic specific toxicity. Compounds containing Cl or Br atoms should not be activating these halogens or be activated by them. Activated Cl or Br can be found in e.g. allylic/propargylic halogenides, activating Cl or Br can be found in e.g. trichloroethanol of pentachlorophenol.

[4]Note that compounds containing benzylic halogens do NOT comply with rule 1.4.1, and thus cannot be considered narcotic chemicals.

[5]Note that compounds containing benzylic halogens do NOT comply with rule 1.4.1, and thus cannot be considered narcotic chemicals. Note also that many of these polycyclic compounds, besides working as narcotics in acute experiments, have chronic toxicities based on specific modes of action.

[6]An example of this would be lindan, or γ-hexachlorocyclohexane, which is much more toxic than the other hexachlorocyclohexanes.

Class 2 type compounds (less inert compounds)

To classify as class 2 type compounds, chemicals should:

be non- or weakly acidic phenols (phenols with one nitro substituent, and/or one to 3 chlorine substituents, and/or alkyl substituents, or

be anilines with one nitro substituent and/or one to 3 chlorine substituents, and/or alkyl substituents, or

be mononitroaromatics with one or two chlorine substituents and/or alkyl substituents, or

be primary alkylamines (containing only C, H, and N), or

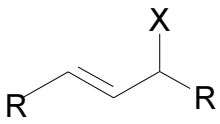
be pyridines with one or two chlorine substituents and/or alkyl substituents.

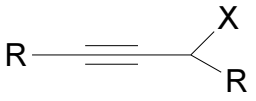
Class 3 type compounds (unspecific reactivity)

In the structures that follow 'R'can be any substituent that consists of C, H, N, O, S and or halogen (not I); 'X' denotes a leaving group. Leaving groups are structures are sufficiently stable (under certain conditions) that they can stabilize an isolated negative charge. Examples of leaving groups are: halogen (Cl, Br, I), cyanide, or hydroxyl group (under acidic or basic conditions).

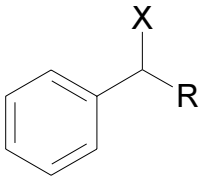
To classify as class 3 type compounds, chemicals should:

possess allylic/propargylic activation. Compounds with a (good) leaving group at an α-position of a carbon-carbon double or triple bond:

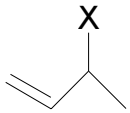


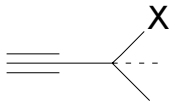


possess benzylic activation. Compounds with a (good) leaving group ar an α-position of an aromatic bond:

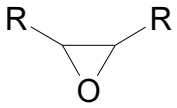


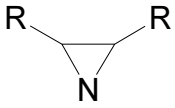
be other compounds with a (good) leaving group at an α-position of a double or triple bond fragment:



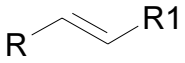


possess a three-membered heterocyclic ring. Compounds containing an epoxide or azaridine function:

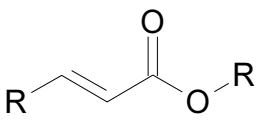


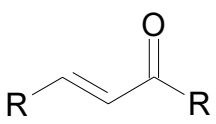


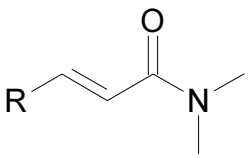
possess activated carbon-carbon double/triple bonds. Compounds containing a polarizable substituent R_1 (carbonyl, nitrile, amide, nitro, sulphone etc.) at an α-position of a duoble of triple bond:

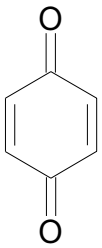


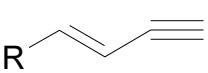
This anables Michael type addition of nycleophiles across the double/triple bond. Examples of such compounds are:

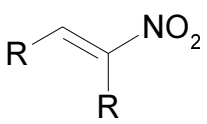


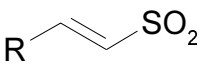




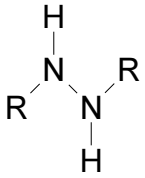
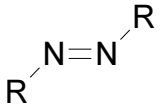
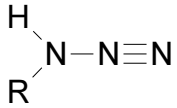




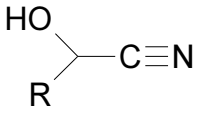
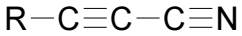




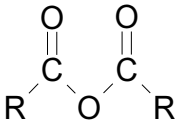
be hydrazines or other compounds with a single, double or triple nitrogen-nitrogen linkage:

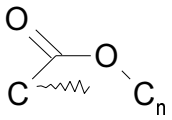




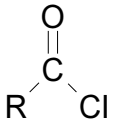
be activated nitriles, like α-hydroxynitriles (cyanohydrins) or allylic/propargylic nitriles:

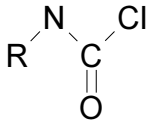



OR contain one of the following structural entities:

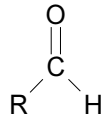
acid anhydrides 

lactones 

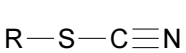
acid halides 

carbamoylhalides 

ketenes

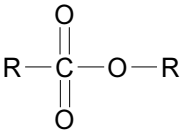
aldehydes 

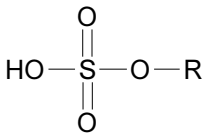
isocyanates

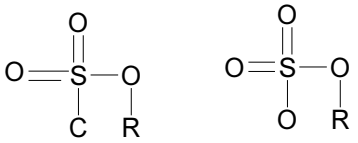
thiocyanates 

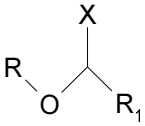
isothiocyanates

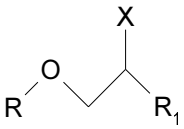
disulphides

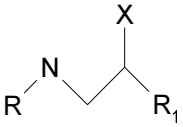
sulphonic ester 

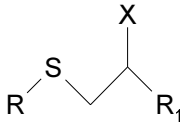
sulphuric esters 

cyclic sulphonic/sulphuric esters 

alpha-haloethers 

beta-haloethers 

nitrogen mustards 

sulphur mustards 

Class 4 (compounds and groups of compounds acting by a specific mechanism)

It is not possible to give definite structural rules for this class. Inclusion in this class must, and should, be based on specific knowledge on mode of toxic action of (groups of) chemicals. Examples of groups of compounds that are known to act by a specific mode of toxic action are:

- DDT and analogues
- (dithio)carbamates
- organotin compounds
- pyrethroids
- organophosphorothionate esters

In order to perform a more rigorous, external validation of the performance of the classification system and the associated RF_Ts, a comparison between estimated aquatic toxicity, based on classification, baseline toxicity estimation and subsequent application of RF_Ts, and actual experimental toxicity data was made for all appropriate chemicals from the ECETOC Aquatic Toxicity database. Estimation of baseline logLC50 was done with the QSAR established for Pimephales promelas:
log LC50 (mol/l) = -0.846 x logKow – 1.385
LogKow values were obtained from the Medchem database.
A total of 176 different compounds were classified according to the classification system as belonging to class 1, class 2, and class 3, 4 combined or not to be classified.
Results: In general the classification system performed as intended, actual fish LC50 values fell within the predicted range. Some compounds were slightly more toxic than predicted by the approach. These invariably are reactive compounds, or compounds amenable to biotransformation, with rather low logKow values and high aqueous solubilities. Some compounds were less toxic than expected, especially so in the high logKow range for class 1 compounds and in the mid-to-high logKow range for class 3 and 4 compounds (Verhaar et al., 2000).

APPENDIX II: Recommendations made in discussion sessions with experts

In brainstorm sessions with some experts various suggestions were made on the procedure of QSAR application for effect assessments. In this appendix these suggestions are presented and commented:

- *If there are QSAR estimates available for the three taxonomic groups, fish, crustaceans and algae, the lowest L(E)C50 is considered concerning the most sensitive species (only if the difference between the highest and the lowest value amounted a factor 10 or more). If such a value for the most sensitive species is available, applying of an assessment factor for inter-species susceptibility is not considered necessary.*

  Although theoretically sound, this proposal has less practical value: [1] with the exception of a chemical group like phenols QSARs on three taxonomic groups or more are not available (see chapter 4.2.2: table 2); [2] in those exceptional cases a large experimental dataset will be available as well, making the use of QSARs needless; [3] typical differences in species sensitivities are to be expected most in chemical groups with specific modes of action for which QSAR are not available.

- *If a QSAR estimate is found to be lower than the reported experimental value, the lower QSAR value should be used for calculation of the MTR (INS). To calculate the HC50 it is recommended to make a decision case by case which value should be used (I-values).*

  No comments; recommendation is adopted.

- *If more than one QSAR estimate is available within a chemical class for species of different taxonomic groups, the lowest value is selected for calculations.*

  No comments; recommendation is adopted.

- *Include the observation by De Zwart (2000), who demonstrated that the average acute toxicity is a factor 10 higher in concentration than the average chronic toxicity based on regression of the acute and chronic alpha values (average of observed L(E)C or NOEC values over a variety of test species) for a large number of chemicals.*

  Recommendation is adopted. Therefore, it is decided to apply a factor 10 for extrapolation of acute toxicity QSAR values to QSAR chronic values.

- *Use the ECOSAR program as the basis for QSAR toxicity estimation.*

  Some QSAR estimates available in ECOSAR were found unreliable. Therefore it is proposed not to use QSAR estimates based on less than 5 experimental data and QSAR estimates with a r^2 beneath 0.7. Applying these restrictions to the ECOSAR database, estimates should not be used for:

- acid chlorides (fish, LC50)
- propargyl alcohols (fish, LC50)
- acrylates (fish 96-h LC50; daphnia,48-h LC50; green algae, 96-h EC50; fish, 32-d ChV)
- aldehydes (fish, 96-h LC50; daphnia, 48-h LC50; fish, 32-d ChV)
- amines, aliphatic (green algae, ChV)
- anilines (fish, 32-d ChV; daphnia, 48-h LC50 and 16-d ChV)
- anilines,amino, meta or 1,3-substituted (fish, 96-h LC50; daphnia, 48-h LC50 and 16-d ChV; green algae, 96-h EC50)
- anilines, amino, ortho or 1,2-substituted (fish, 96-h LC50; daphnia, 48-h LC50; green algae, 96-h EC50)
- anilines, amino, para of 1,4-substituted (fish, 96-h LC50; daphnia 48-h LC50; green algae, 96-h EC50)
- anilines, dinitro (fish, 96-h LC50and 32-d ChV; daphnia, 48-h LC50)
- aziridines (fish acute LC50; daphnia 48-h LC50; green algae 7-d ChV)
- benzenes, dinitro (fish, 96-h LC50 and 32-d ChV; daphnia, 48-h LC50 and 16-d ChV)
- benzotriazoles (fish, 96-h LC50; daphnia 48-h LC50 uses the SAR for neutral organics; green algae, 96-h EC50)
- carbamates (sea urchin 48-h NOEC)
- carbamates, dithio (SARs are sigmoidal; statistically poor relationships)
- crown ethers, use SARs for neural organics?
- diazoniums, aromatic (fish, 96-h LC50)
- epoxides, mono (fish, 96-h LC50; daphnia 48-h LC50)
- epoxides, di (fish, 96-h LC50and 14-d LC50; daphnia 48-h LC50)
- esters (daphnia 48-h LC50, use SAR neutral organics; green algae, 96- h EC50 and 16-d ChV)
- esters, mono (fish, 32-d ChV)
- esters, di, aliphatic (fish, 32-d ChV)
- esters phosphate (fish, 96-h LC50)
- esters phthalate (fish, 96-h LC50; daphnia 48-h LC50; use ester SAR for acte toxicityfor fish; daphnia 21-d NOEC, use netral organic SAR for following phthalate esters: aliphatic diesters, aromatic diesters, aliphatic-aromatic diesters, and phthalates derived from aliphatic alcohols and phenol)
- hydrazines (daphnia, 48-h LC50; green algae, 144-h EC50)
- imides (fish, 96-h LC50)
- ketones, di, aliphatic (daphnia, 16 d ChV; green algae, ChV)
- malononitriles (fish, 96-h LC50)
- neutral organics (earthworm 14-d LC50)
- peroxide acids (fish, 96-h LC50; daphnia 48-h LC50)
- phenols (daphnia 48-h LC50; green algae, 96-h EC50 and ChV, use SARs for neutral organics; fish 60-d ChV)
- phenols, dinitro (fish, 96-h LC50 and 32-d ChV; daphnia, 16-d ChV)
- polymers, polycationic
- surfactants, anionic (fish 96-h LC50 and 28-d NOEC; daphnia, 48-h LC50 and 21-d NOEC)
- surfactants, cationic, quaternary ammonium, monoalkyl, see Clements et al. 1996
- surfactants, cationic, quaternary ammonium, dialkyl (daphnia, 48-h LC50; green algae, 96 h EC50 and ChV)
- surfactants, nonionic
- surfactants, ethomeen
- thiazolinones, iso (fish, 96-h LC50; daphnia, 48-h LC50; green algae, 96-h EC50 and ChV
- thiols and mercaptanes (fish 96-h LC50; daphnia 48-h LC50
- triazines, substituted, use for fish and daphnia SARs for neutral organics

- *The ECOSAR system is a compilation of functional group QSARs; if the chemical in question has more than one functional groups, ECOSAR gives QSAR estimates for all of these groups and the user of the program has to choose wich estimate is the most appropriate one to use.*

  In this case it is proposed to use the QSAR that generates to lowest value.

- *ECOSAR classifies the chemicals automatically into one of the 42 chemical classes that are distinguished in ECOSAR; this forms a handy tool to classify the chemicals according to Verhaar et al. (1992), which is often difficult and complex. See table below for the integration of the ECOSAR chemical classes.*

  No comments; recommendation is adopted.

- *In view of the higher predictive power of the computerized model CATALYST and the CoMFA model compared to the classical approach with logKow value as sole descriptor, it is recommended to gather more detailed information on these programs. Investigations must be made about the possibilities and the required knowledge to use such models within RIVM/CSR in future.*

  No comments; recommendation is adopted.

- *Combine the recent observations by De Zwart (2000) on the relationship between the standard deviation or the slope, and the toxic mode of action with the former proposal of Verhaar et al. (1992; see figure 4.3, section 4.2.1) to improve the effect assessment.*

  Apart from classifying chemicals according to their mode of action, Verhaar et al. (1992) found that, in general, effect concentrations for less inert chemicals (class 2) are between 5 and 10 times lower than predicted by baseline toxicity QSAR equation, whereas effect concentrations for reactive chemicals (class 3), as well as for specifically acting chemicals (class 4) are between 10 and 10^4 times lower than predicted by baseline toxicity QSAR equation. These observations can be used to define toxicity range factors RF_Ts, multiplication factors to estimate baseline toxicity of a chemical, as presented in figure 4.3.

  De Zwart (2000) obtained for a large number of both organic and inorganic toxicants, log-ogistic species sensitivity distributions curves (SSDs). The log-logistic sensitivity model is characterized by two parameters: (alpha) which is the average of the observed log tranformed L(E)C or NOEC values over a variety of test species, and (beta) which is proportional to the standard deviation of the log transformed toxicity values. De Zwart found that the beta values are related to the toxic mode of action of the compound under consideration and therefore these beta values could be used to estimate the HC5 with the following equation:

$$\frac{1}{1+e^{\frac{-x+\alpha}{\beta}}} = 0.05$$

  In the scheme below it is shown how the different chemical classes used in ECOSAR and the different modes of toxic action used by de Zwart (2000) can be integrated in the classification scheme of Verhaar et al. (1992). In ECOSAR reliable QSARs are available for the chemical groups printed bold.

Scheme I: Integration of the different chemical classes used in ECOSAR and the different mode of action used by De Zwart (2000) in the chemical classification scheme of Verhaar (1992).

Class 1	Class 2	Class 3	Class 4	Verhaar et al, 1992
neutral organics	**aliphatic amines** **anilines** subst. anilines **phenols**	acid chlorides acrylates, **methacrylates** aldehydes aziridines benzotriazoles dinitroanilines dinitrobenzenes **dinitriphenols** diazoniums epoxides **esters** **hydrazines** imides **ketones** malonitriles peroxy acids thiols	Carbamates crown ethers polymers surfactants thiazolinones triazines **ureas**	ECOSAR
Non polar narcosis	polar narcosis	Diesters	AchE inhibition: Organophosphates Carbamates Photosynthesis Inhibitors amino acid synthesis inh. uncoupler oxid. phosphoryl. multi-site inhib. systemic fungic. systemic herbic. sporulation inh. gemination inh. Neurotoxicants: Pyrethroids cyclodiene type DDT-type plant growth inh. and regulator cell division inh. Quinolines	de Zwart, 2000

An overview on the (acute) beta values of De Zwart (2000) is given in table 1, including the classification of the chemical groups according to Verhaar et al. (1992). This table shows that most beta values are derived on the basis of a great amount of experimental data on chemicals with a specific mode of toxic action in most cases. However, for these classes of chemicals QSARs are scarce. Therefore, the benefit of using these beta values in toxicity estimation is not considered to be of great importance.
In the proposal of Verhaar et al. (1992) Toxicity Range Factors (RF_Ts) are applied to derive a toxicity estimation. If the substance considered, is a reactive chemical or has a specific mode of action (class 3 or 4) the RF_Ts are very high (10^4) and may overestimate the toxicity. As such the result eliminates the advantage of QSAR application in deriving effect based intervention values in case only one acute toxicity data is available and an extrapolation factor of 1,000 is applied.

Table 1 *Acute beta values, based on 10 or more species, averaged over toxic modes of action (n = number of compounds; SEM = standard error of the mean).*

Toxic Mode of Action	n	Avg. Beta	SEM	classified acc. Verhaar
Non polar narcosis	34	0.39	0.03	1
Acetylcholinesterase inhibition: organophosphates	27	0.71	0.03	4
Inhibits photosynthesis	20	0.60	0.03	4
Polar narcosis	13	0.31	0.03	2
Acetylcholinesterase inhibition: carbamates	11	0.50	0.05	4
Uncoupler of oxidative phosphorylation	8	0.38	0.05	4
Multi-site inhibition	6	0.62	0.07	4
Dithiocarbamates	6	0.57	0.05	4
Diesters	6	0.42	0.07	3
Systemic fungicide	5	0.46	0.04	4
Sporulation inhibition	5	0.37	0.05	4
Neurotoxicant: pyrethroids	4	0.65	0.03	4
Neurotoxicant: cyclodiene-type	4	0.61	0.01	4
Plant growth inhibition	4	0.52	0.06	4
Membrane damage by superoxide formation	3	0.69	0.01	4
Cell division inhibition	3	0.63	0.21	4

APPENDIX III: Example of an ECOSAR record

SMILES : O=C
CHEM : Formaldehyde
CAS Num: 000050-00-0
ChemID1:
ChemID2:
ChemID3:
MOL FOR: C1 H2 O1
MOL WT : 30.03
Log Kow: 0.35 (KowWin estimate)
Melt Pt:
Wat Sol: 6434 mg/L (calculated)

ECOSAR v0.99f Class(es) Found

Aldehydes

ECOSAR Class	Organism	Predicted Duration	End Pt	mg/L (ppm)
Neutral Organic SAR (Baseline Toxicity)	: Fish	14-day	LC50	1103.201
Aldehydes	: Fish	96-hr	LC50	8.297
Aldehydes	: Daphnid	48-hr	LC50	16.071
Aldehydes	: Green Algae	96-hr	EC50	430.284
Aldehydes	: Fish	32-day	ChV	2.688
Aldehydes	: Green Algae		ChV	16.581

Note: * = asterick designates: Chemical may not be soluble enough to measure this predicted effect.
Fish and daphnid acute toxicity log Kow cutoff: 6.0
Green algal EC50 toxicity log Kow cutoff: 6.4
Chronic toxicity log Kow cutoff: 7.0
MW cutoff: 1000

APPENDIX IV: Example of an EPIWIN estimation record

SMILES : Oc1cc(CL)c(CL)c(CL)c1CL
CHEM : 2,3,4,5-Tetrachlorophenol
CAS NUM: 004901-51-3
MOL FOR: C6 H2 CL4 O1
MOL WT : 231.89
---------- EPI SUMMARY (v2.30) ----------
Physical Property Inputs:
Water Solubility (mg/L): ------
Vapor Pressure (mm Hg) : ------
Henry LC (atm-m3/mole) : ------
Log Kow (octanol-water): ------
Boiling Point (deg C) : ------
Melting Point (deg C) : ------

Log Octanol-Water Partition Coef (SRC):
Log Kow (KOWWIN v1.57 estimate) = 4.09
Log Kow (Exper. database match) = 4.21

Boiling Pt, Melting Pt, Vapor Pressure Estimations (MPBPWIN v1.27):
Boiling Pt (deg C): 288.07 (Adapted Stein & Brown method)
Melting Pt (deg C): 85.80 (Mean or Weighted MP)
VP(mm Hg,25 deg C): 0.000339 (Modified Grain method)

Water Solubility Estimate from Log Kow (WSKOW v1.27):
Water Solubility at 25 deg C (mg/L): 28.69
log Kow used: 4.21 (expkow database)
no-melting pt equation used

Henrys Law Constant (25 deg C) [HENRYWIN v3.00]:
Bond Method : 1.69E-007 atm-m3/mole
Group Method: 3.46E-007 atm-m3/mole

Probability of Rapid Biodegradation (BIOWIN v2.62):
Linear Model : 0.0233
Non-Linear Model : 0.0006
Expert Survey Biodegradation Results:
Ultimate Survey Model: 1.9167 (months)
Primary Survey Model : 2.8915 (weeks)

Atmospheric Oxidation (25 deg C) [AopWin v1.85]:
Hydroxyl Radicals Reaction:
OVERALL OH Rate Constant = 1.5459 E-12 cm3/molecule-sec
Half-Life = 6.919 Days (12-hr day; 1.5E6 OH/cm3)
Half-Life = 83.026 Hrs
Ozone Reaction:
No Ozone Reaction Estimation
Reaction With Nitrate Radicals May Be Important!

Soil Adsorption Coefficient (PCKOCWIN v1.62):
Koc : 2002
Log Koc: 3.302

Aqueous Base/Acid-Catalyzed Hydrolysis (25 deg C) [HYDROWIN v1.63]:
Rate constants can NOT be estimated for this structure!

BCF Estimate from Log Kow (BCFWIN v2.0):
Log BCF = 2.142 (BCF = 138.6)
log Kow used: 4.21 (expkow database)

Volatilization from Water:
Henry LC: 3.46E-007 atm-m3/mole (estimated by Group SAR Method)
Half-Life from Model River: 3870 hours (161.2 days)
Half-Life from Model Lake : 2.826E+004 hours (1177 days)

Removal In Wastewater Treatment:
Total removal: 40.47 percent
Total biodegradation: 0.40 percent
Total sludge adsorption: 40.06 percent
Total to Air: 0.01 percent

APPENDIX V: Mailing list

1. Directie RIVM
2. Depot Nederlandse Publicaties en Nederlandse Bibliografieën
3. Sectordirecteur Milieuonderzoek
4. Sectordirecteur Stoffen en Risico's
5. Hoofd Centrum voor Stoffen en Risicobeoordeling
6. dr. J. Hermens (RITOX)
7. dr. W. Peijnenburg (RIVM/ECO)
8. drs. D. de Zwart (RIVM/ECO)
9. prof. dr. C. van Leeuwen (RIVM/CSR)
10. mw. dr. ir M. Pieters (RIVM/CSR)
11. drs. R. Luttik (RIVM/CSR)
12. drs. T. Vermeire (RIVM/CSR)
13. mw. drs. E. Hulzebos (RIVM/CSR)
14. mw. L. Wijkhuizen-Maslankiewicz (RIVM/CSR)
15. dr. D. Sijm (RIVM/CSR)
16. mw. dr. A. van Wezel (RIVM/CSR)
17. dr. E. Verbruggen (RIVM/CSR)
18. drs. T. Traas (RIVM/CSR)
19. dr. W. Slooff (RIVM/CSR)
20. mw. R. Posthumus (RIVM/CSR)
21. SBD/Voorlichting & Public Relations
22. Bureau Rapportenregistratie
23. Bibliotheek RIVM

24-38 Bureau Rapportenbeheer

39-60 Reserve exemplaren