J. Braz. Chem. Soc., Vol. 13, No. 6, 754-762, 2002.
Printed in Brazil - ©2002 Sociedade Brasileira de Química
0103 - 5053 $6.00+0.00

Review

Prediction of Environmental Toxicity and Fate Using Quantitative Structure-Activity Relationships (QSARs)

John C. Dearden

School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, England

Atualmente, muito pouco se sabe sobre a toxidez de mais de 100.000 substâncias químicas que são liberadas no meio-ambiente. O custo necessário para a obtenção dessas informações é elevado não apenas em termos financeiros, mas também no consumo de tempo e de animais. Por isso, muitas indústrias e agências governamentais regulatórias estão focalizando sua atenção na predição da toxidez e suas causas através de relações entre a estrutura química e a atividade (QSARs). Este artigo examina o uso de QSARs neste contexto. Como, em geral, as QSARs são dependentes de mecanismos específicos, a primeira etapa é classificar a substância tóxica em uma das quatro classes de toxidez: narcose apolar, polar, reatividade inespecífica e ação específica (por exemplo, atividades anticolinesterase). Uma QSAR apropriada pode, então, ser selecionada para predizer a toxidez de uma dada substância química. Também existem sistemas inteligentes para a predição da toxidez. As predições de bioconcentrações, absorções do solo e biodegrabilidade também podem ser realizadas. A predição por QSARs e sistemas inteligentes de propriedades físico-químicas, tais como o coeficiente de partição, solubilidade aquosa, pontos de fusão e ebulição, pressão de vapor e constante de lei de Henry, pode ser prontamente obtida.

Little or nothing is known about the toxicity of most of the >100,000 chemicals released into the environment. The cost of obtaining such information experimentally would be enormous in terms of money, time and animals. Companies and regulatory agencies are therefore turning to the prediction of environmental toxicity and fate through the use of quantitative structure-activity relationships (QSARs). This paper examines the use of QSARs in this context. Generally, QSARs are mechanism-specific, so that the first step is to classify a toxicant into one of four broad classes of toxicity: non-polar narcosis, polar narcosis, unselective reactivity and specific action (*e.g.* anticholinesterase activity). An appropriate QSAR can then be selected in order to predict the toxicity of a given chemical. There are also some expert systems available for toxicity prediction. QSAR predictions of bioconcentration, soil sorption and biodegradability can also be made; again, expert systems are available for such prediction. QSAR and expert system prediction of physico-chemical properties such as partition coefficient, aqueous solubility, melting and boiling point, vapour pressure and Henry's law constant can readily be made.

Keywords: QSAR, environmental toxicity, environmental fate, physicochemical properties

1. Introduction

Over 100,000 chemicals are released into the environment, and as few as 1 – 5% have toxicity data available. Even for the high production volume chemicals or HPVCs (those chemicals produced in quantities of > 1000 tonnes per year in the EU or > 1,000,000 pounds (about 442 tonnes) per year in the U.S.A.) there is a paucity of information concerning their toxicity,[1] as Table 1 shows.

With increasing concern about the environment, governments and regulatory agencies worldwide are seeking to assess the ecotoxicological risks posed by the release of chemicals. For example, in February 2001 the Commission of the European Communities presented a White Paper entitled "Strategy for a Future Chemicals

Table 1. Availability of toxicity data for high production volume chemicals

	European Union	U.S.A.
Number	2465	2863
Full toxicity data	3%	7%
Partial toxicity data	43%	50%
No toxicity data	54%	43%

* e-mail: j.c.dearden@livjm.ac.uk

Policy" which proposed that some 30,000 existing chemicals be tested on animals for a range of toxic effects.[2] This would clearly be an extremely expensive and time-consuming undertaking, involving the use of many thousands of animals. Table 2 gives information[3] showing the costs of some of the tests.

Table 2. Costs of single toxicity tests in 1994

Test	Cost (US$)	Test duration
Microtox	62	5 min
Lettuce	296	96 h
Rotifer	333	24 h
Brine shrimp	333	24 h
Polytox	407	21 min
Daphnia magna	703	48 h
Pimephales promelas	1036	96 h
Selenastrum capricornutum	1280	96 h
Carcinogenicity (rodent)	2 million	2 year

One also has to query whether it is necessary to test the many chemicals that have been in use for many years without obvious adverse effects.

In view of the increasing demands for toxicity assessment, a number of organisations are investigating the use of alternatives to animal testing. For example, ECVAM (European Centre for the Validation of Alternative Methods; altweb.jhsph.edu/publications/ECVAM/ecvam-reports.htm) is examining ways forward for the application of existing alternative methods, and seeking to identify areas of research that could facilitate the development of new alternative methods; ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals; www.ecetoc.org) recently (March 2002) organised a Workshop on Regulatory Acceptance of (Q)SARs for Human Health and Environmental Endpoints; FRAME (Fund for Replacement of Animals in Medical Experimentation; www.frame.org.uk) is committed to the development and acceptance of methods to replace animal testing for regulatory and other purposes.

2. Methods

2.1 Quantitative structure-activity relationships

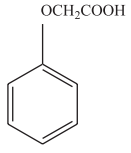
One of the chief alternatives to animal testing for toxicity is the use of quantitative structure-activity relationships (QSARs), which are mathematically-derived rules that quantitatively describe a property in terms of descriptors of chemical structure. Of course, biological data are needed to develop a QSAR in the first place, but it can then be used to predict the toxicities of other chemicals with the same mechanism of action. It should also be emphasised that predictions should not be made on chemicals that are outside the range covered by the chemicals used to develop the QSAR (the training set).

It is often difficult to determine whether or not a chemical possesses a particular mechanism of action. For this reason QSARs are usually developed using compounds of a single chemical class (*e.g.* phenols) on the assumption that such a congeneric series has a common mechanism of action. Any chemicals that do not possess the same mechanism of action will show up as outliers; that is, they will not be well modelled by the QSAR.

The descriptors used in the development of a QSAR are physico-chemical and structural properties. They fall into three broad classes – hydrophobic, electronic and steric.[4] Most chemicals move through an organism by a partitioning process between aqueous and lipid compartments, so that transport is controlled largely by hydrophobicity. This is generally well modelled by the octanol-water partition coefficient (P). Interaction with a receptor site, on the other hand, is a function of the ability of the chemical to form (generally reversible) links with the receptor (through hydrogen bonding and dipolar forces, for example) and by the ability to fit the receptor site well, which is a function of molecular size and shape. Dearden[4] has discussed each class of descriptor in detail.

Topological descriptors are also widely used in QSAR. These are derived from the molecular structure, and are not always easy to interpret in physico-chemical terms. The most extensively used topological descriptors are molecular connectivities[5] and electrotopological state descriptors.[6]

Corwin Hansch is acknowledged to be the father of the discipline of QSAR; his first publication on the subject[7] dealt with the herbicidal effects of derivatives of phenoxyacetic acid (**1**). He has since published many hundreds of papers, and his recent book[8] summarises progress in the field.



1

2.2 QSAR in environmental toxicology

QSAR is a tool for the prediction of biological activity, and thus lends itself readily to the prediction of

environmental toxicity. Over the past 20 years environmental QSAR has increased steadily in importance, and Nendza[9] has admirably summarised its achievements. It has now reached the stage where some regulatory agencies, such as the U.S. Environmental Protection Agency, routinely use some QSAR–predicted toxicities for regulatory purposes; it is anticipated that such use will increase greatly in the future, as more assurances are sought on the safety of chemicals, and more public pressure is brought to bear against the use of animals in toxicity testing. It should be noted, however, that experimental toxicity data are needed in the first place in order to develop a QSAR, and there is still a shortage of good quality data in many areas.

2.3 QSARs for toxicity prediction

Most environmental toxicity data have been obtained using aquatic animals such as fish of various species, *Daphnia*, *Tetrahymena pyriformis*, *Vibrio fischeri* and algæ. Cronin and Dearden[10] have reviewed the literature concerning QSAR modelling of aquatic toxicity. Eight modes of action have been identified in fish, namely non-polar narcosis, polar narcosis, uncoupling of oxidative phosphorylation, respiratory membrane irritation, acetylcholinesterase inhibition, central nervous system seizure, inhibition of photosynthesis, and alkylation.[11] However, these are generally more broadly grouped as: non-polar narcosis, polar narcosis, unselective reactivity, and specific mechanisms of action. It is important, in order to obtain a correct QSAR prediction of toxicity, that a chemical's mode of action is correctly identified. To this end Verhaar *et al.*[12] developed a scheme based on the presence of functional groups to classify chemicals into these four groups. Later Boxall *et al.*[13] used a pattern recognition approach to four-group classification based on 7 molecular descriptors, and obtained 76% correct predictions.

The first QSAR correlation of non-polar narcosis was developed by Könemann,[14] who correlated the acute toxicity of diverse industrial chemicals to the guppy, *Poecilia reticulata*:

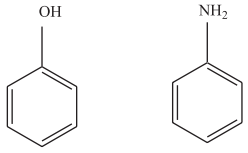
$$\log 1/LC_{50} = 0.87 \log P - 1.87 \quad (1)$$
$$n = 50,\ r^2 = 0.98,\ s = 0.23$$

where LC_{50} = concentration (mmol L^{-1}) to kill 50% of fish in a specified time, n = number of chemicals, r = correlation coefficient, and s = standard error of the estimate.

van Leeuwen *et al.*[15] later showed that similar correlations obtained for other aquatic species. Lipnick[16] proposed that non-polar narcosis be considered as "base-line" toxicity, with no chemicals having lower toxicity, and this is now accepted. It should be pointed out that occasionally a chemical appears to show lower than base-line toxicity, but this is invariably due to an artefact, such as evaporative loss giving a lower than nominal aqueous concentration.

Veith and Broderius[17] reported that some unreactive chemicals, such as phenol (**2**) and aniline (**3**) derivatives, that produced toxicity consistent with narcosis were nevertheless more toxic than would be expected for non-polar narcosis. This mode of action is now termed polar narcosis, and may result from the presence of a strongly hydrogen bonding group in a molecule. The correlation found by Veith and Broderius[17] for toxicity to the fathead minnow, *Pimephales promelas*, was:

$$\log 1/LC_{50} = 0.65 \log P - 0.71 \quad (2)$$
$$n = 39,\ r^2 = 0.900,\ s \text{ not given}$$



2 **3**

Reactive chemicals are more generally toxic still, although their toxicities can nevertheless sometimes be correlated with log P alone. An example is given by the toxicity of α,β-unsaturated aldehydes such as 2-hexenal (**4**) to a phosphorescent bacterium, *Vibrio fischeri*:[18]

$CH_3CH_2CH_2CH{=}CH.CHO$
4

$$\log 1/EC_{50} = 0.50 \log P + 0.35 \quad (3)$$
$$n = 7,\ r^2 = 0.854,\ s = 0.23,\ F = 36.2$$

It can be seen that the coefficients on log P are in the order: equation 1 > equation 2 > equation 3, whilst the opposite is true of the intercepts. This means that at some high value of log P, the three equations converge, as is shown in Figure 1.

Clearly, correlations such as those depicted in Figure 1 do not extend *ad infinitum*. As hydrophobicity increases, aqueous solubility decreases, and a point is reached where solubility is too low for a toxic concentration to be reached (the solubility cut-off). This is typically in the region of log P ~ 6-7.

Generally, the toxicity of reactive chemicals can be modelled only by the inclusion of one or more descriptors that reflect reactivity. Typically such reactivity is

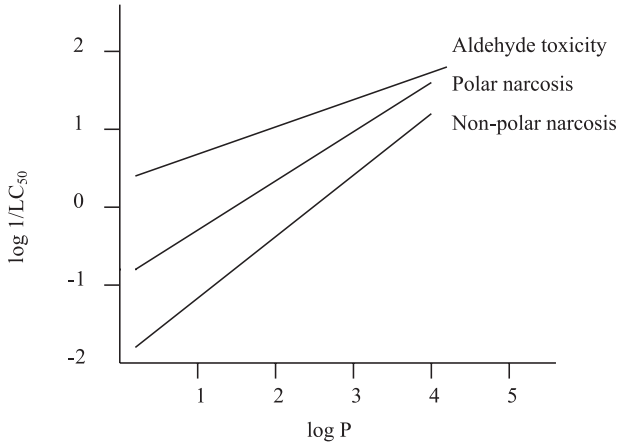


Figure 1. QSARs for non-polar narcosis (equation 1), polar narcosis (equation 2) and aldehyde toxicity (equation 3).

electrophilic in nature, since nucleophilic groups such as NH, OH and SH abound in biological macromolecules. Cronin and Schultz[19] found the following QSAR for the toxicity of aromatic compounds to an aquatic ciliate, *Tetrahymena pyriformis*:

$$\log 1/IGC_{50} = 0.603 \log P - 0.330 E_{LUMO} - 1.00 \quad (4)$$
$$n = 239, r^2 = 0.800, Q^2 = 0.796, s = 0.335, F = 476$$

where IGC_{50} = the concentration (mmol L^{-1}) to inhibit growth by 50%, E_{LUMO} = energy of lowest unoccupied molecular orbital, Q = cross-validated correlation coefficient (leave-one-out procedure), and *F* = Fisher statistic.

The statistics given for equation 4 are the preferred statistics for the reporting of QSARs; the cross-validated correlation coefficient is a measure of the predictivity of the QSAR (as distinct from its merely correlative ability); *F* gives a measure of probability that the correlation has not arisen by chance.

Since most QSARs are developed for predictive purposes, it is important that their predictive ability is assessed. Whilst cross-validation is one method of assessing predictive ability, it is an internal method, and can be criticised as simply giving an indication of the internal consistency of the data-set used to develop the QSAR (the training set). A better method is to use the QSAR to predict the toxicities of chemicals not used in the training set (external validation).

As well as E_{LUMO}, other measures of reactivity can be used in QSARs involving reactive chemicals. An interesting example is given by Mekenyan and Veith[20], who correlated the toxicity of chemicals with three different modes of action to *P. promelas* by the use of the molecular orbital descriptor known as superdelocalisability:

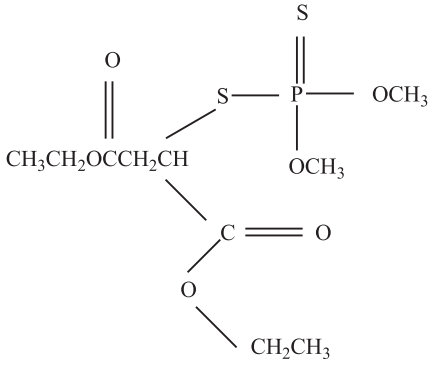
$$\log 1/LC_{50} = 0.56 \log P + 13.7 S_{av}^{N} - 1.49 \quad (5)$$
$$n = 114, r^2 = 0.81, s \text{ not given}$$

where S_{av}^{N} = average nucleophilic superdelocalisability.

Specifically-acting chemicals are typified by the organophosphorus insecticides such as malathion (**5**). Hermens *et al.*[21] found that they could model the toxicity of these compounds to *P. reticulata* with a two-term QSAR:

$$\log 1/LC_{50} = 0.23 \Sigma\pi + 0.80 k + 2.77 \quad (6)$$
$$n = 9, r^2 = 0.92, s = 0.19$$

where $\Sigma\pi$ = sum of hydrophobic fragmental constants, and k = experimentally determined reaction rate constant with 4-nitrobenzylpyridine. A later study by de Bruijn and Hermens[22] with 20 such insecticides failed to confirm this correlation. However, Verhaar *et al.*[23], using a number of physico-chemical descriptors, obtained a series of good QSARs for the toxicity of a series of 12 organophosphorus insecticides. Recently Devillers[24] used a neural network approach to model the toxicity of 66 diverse pesticides to the bluegill fish, *Lepomis macrochirus*, using autocorrelation descriptors encoding lipophilicity and hydrogen bonding ability; he found RMS residual values of 0.345 log units for the training set and 0.359 log units for a 21-compound test set.



5

An interesting study by Kaiser *et al.*[25] used probabilistic neural networks to model the toxicity to *P. promelas* of a very diverse data-set of 1000 chemicals with various modes of action. Using functional groups as descriptors, they obtained $r^2 = 0.899$, and a test set of 84 chemicals yielded $r^2 = 0.803$.

2.4 Expert systems for toxicity prediction

An expert system has been defined[26] as "any formalised system, not necessarily computer-based, which enables a

user to obtain rational predictions about the properties of chemicals". Greene[27] has recently reviewed expert systems for toxicity prediction.

A number of such systems are available commercially, and whilst most of them deal predominantly with human health hazards such as carcinogenicity, teratogenicity and skin sensitisation, some include modules for the prediction of ecotoxicological endpoints. TOPKAT (www.accelrys.com) uses a QSAR approach, based largely on topological descriptors, to predict LC_{50} values for *Daphnia magna* and *P. promelas*. MULTICASE (www.multicase.com) identifies structural fragments that are responsible for a given toxicity, and includes fish LC_{50} among available endpoints. HAZARDEXPERT (www.compudrug.com) incorporates expert human knowledge to identify structural fragments linked to toxicity; it provides predictions across a range of trophic levels with different dosing regimes. An associated program, METABOLEXPERT, predicts metabolites that can then also be assessed by HAZARDEXPERT. ASTER (www.epa.gov/med/databases/aster.html) was developed by the U.S. Environmental Protection Agency, and the endpoints covered are largely ecotoxicological, namely LC_{50} values for *P. promelas*, sheepshead minnow (*Cyprinodon variegatus*) and *D. magna.* It is essentially a database, but uses QSAR to predict toxicities when experimental data are not available. It uses different QSAR models for different types of chemicals (non-polar narcotics, polar narcotics and so on), as is shown to be appropriate by Figure 1. Owing to security restrictions, ASTER is currently not publicly available. Finally, the U.S. E.P.A. has also developed ECOSAR (www.epa.gov/oppt/exposure/docs/episuitedl.htm), which is freely downloadable from the website. It uses hydrophobicity-based QSARs to predict toxicities to fish, daphnids and green algae.

2.5 QSARs for bioconcentration

The accumulation of chemicals in biota from the environment represents a considerable hazard. There are two main routes of uptake: *via* the food chain, thus producing higher concentrations in higher trophic levels, termed bioaccumulation; and uptake from the surrounding mileu, termed bioconcentration. The former has not been subjected to QSAR analysis, important though it is. Bioconcentration has, however, been extensively investigated in this way. Typically, an organism takes up a toxicant from a surrounding aqueous phase, which may be regarded as a partitioning process. It follows that bioconcentration should be related to log P, and this is indeed the case. Nendza[9] has comprehensively reviewed QSAR modelling of bioconcentration.

Numerous bioconcentration QSARs have been published, and the following two examples, both involving diverse chemicals, will suffice to illustrate. Mackay,[28] using fish bioconcentration data, developed the following QSAR:

$$\log BCF = 1.00 \log P - 1.32 \quad (7)$$
$$n = 44,\ r^2 = 0.95,\ s = 0.25$$

where BCF = bioconcentration factor, the ratio of concentration in the fish to that in the surrounding aqueous phase.

Geyer *et al.*[29], using algal bioconcentration data, reported a similar dependence on hydrophobicity:

$$\log BCF = 0.681 \log P + 0.164 \quad (8)$$
$$n = 41,\ r^2 = 0.814,\ s \text{ not given}$$

However, it has been observed that such rectilinear correlations break down at high log P values (> 6-7). There are several possible reasons for this: very hydrophobic chemicals may not have reached equilibrium in the organism during the test; very hydrophobic molecules are often very large, and large (MW > 500) molecules have great difficulty in penetrating membranes; hydrophobic chemicals tend to metabolise more quickly[30], thus reducing the concentration of the original toxicant in the organism; for large molecules, octanol may not be a good surrogate for lipid; certain specific sub-structural effects, such as in 2,4-dinitrophenols, appear to reduce bioconcentration.[31]

QSARs that are biphasic in log P have been developed to try to model this non-rectilinear behaviour. For example, Dimitrov *et al.*,[32] using a large diverse data-set for fish BCF, developed an unusual QSAR

$$\log BCF = 0.420 + 3.321\ e^{-(\log P - \log P_0)^2/10.15} \quad (9)$$
$$n = 443,\ r^2 = 0.73,\ s = 0.65$$

where $\log P_0$ = optimal log P (in this case 6.35).

Sabljić[33] used second-order valence molecular connectivity to model the bioconcentration of a diverse group of chemicals in fish:

$$\log BCF = 2.12\ {}^2\chi^v - 0.16\ ({}^2\chi^v)^2 - 2.13 \quad (10)$$
$$n = 84,\ r^2 = 0.933,\ s = 0.345$$

The rationale behind the use of molecular connectivity is not clear. ${}^2\chi^v$ is known to correlate with molecular size,[34] and, as mentioned earlier, large molecules do not penetrate lipid membranes readily; it may be also that, in the chemicals

used in this study, hydrophobicity and size were reasonably collinear. Equation 10 should not, however, be construed as indicating that bioconcentration is largely a function of molecular size. In fact, as equations 7 and 8 show, log BCF is a rectilinear function of log P up to log P values of 6-7.

There is one expert system available for the prediction of bioconcentration, namely BCFWIN, developed by Syracuse Research Corporation and freely downloadable from the E.P.A. website (www.epa.gov/oppt/exposure/docs/episuitedl.htm).

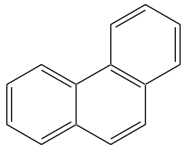
2.6 QSARs for soil sorption

The sorption of chemicals to soil and sediment is an important factor in their distribution and mobility. The extent of sorption of a chemical is, of course, a function of its molecular structure, but depends also on such soil factors as particle size, porosity, pH and organic carbon content. Indeed, concerning the last-mentioned, it is generally accepted that little or no sorption occurs to silica, and that sorption is directly proportional to organic carbon content of the soil. Nendza[9] has reviewed in detail the QSAR analysis of soil sorption.

The sorption coefficient K is defined as (concentration of chemical sorbed to soil) ÷ (concentration of chemical in surrounding aqueous phase). It is generally referred to as K_{oc} (OC = organic carbon). Since effectively the chemical partitions between the surface of the soil and the aqueous phase, it is not surprising that K_{oc} correlates with log P, and many sorption QSARs have confirmed this.[9] Two examples illustrate this. Briggs[35] found an excellent correlation for a large group of pesticides:

$$\log K_{oc} = 0.52 \log P + 1.12 \quad (11)$$
$$n = 105,\ r^2 = 0.90,\ s \text{ not given}$$

For a series of aromatics and polyaromatic hydrocarbons (PAHs, **6**), Hodson and Williams[36] found:



6
Phenanthrene, an example of a PAH

$$\log K_{oc} = 0.83 \log P + 0.29 \quad (12)$$
$$n = 20,\ r^2 = 0.90,\ s \text{ not given}$$

A number of workers have correlated K_{oc} values with descriptors that effectively model molecular size, such as molar refractivity and first- and second-order molecular connectivities. An example is the study by Sabljić[37] of PAHs and halogenated hydrocarbons:

$$\log K_{oc} = 0.55\ {}^{1}\chi + 0.45 \quad (13)$$
$$n = 37,\ r^2 = 0.95,\ s = 0.34$$

It is likely that such descriptors are simply reflecting the known collinearity between hydrophobicity and molecular size within congeneric series, for, as Nendza[9] has pointed out, there is no unique dependence of soil sorption on the structural features encoded in these descriptors.

It is interesting to note that no-one appears to have reported QSAR correlations of K_{oc} values for diverse groups of chemicals. The organic content of soils must, by virtue of the wide range of chemicals that can be sorbed, and the fact that K_{oc} is directly correlated with hydrophobicity, act as a non-specific binding site. There does not therefore seem to be any reason why one could not correlate K_{oc} values of a diverse group of chemicals with hydrophobicity. It should be borne in mind, however, that in general real soils are used for sorption, and thus each data-set in effect uses a different protocol. This probably explains the very wide range of slopes and intercepts observed in soil sorption QSARs; in the nine log K_{oc}-log P correlations listed by Nendza,[9] slopes range from 0.38 to 0.99, and intercepts from – 0.35 to + 1.92. It would be useful to know whether or not K_{oc} values of diverse chemicals, measured under a single protocol and using a single type of soil, would correlate with hydrophobicity. If so, that would greatly enhance the ability to predict K_{oc} values; even then, the results would probably not be applicable to other soil types.

For the present, however, it must be accepted that K_{oc} predictions are valid only within specific chemical classes for which QSARs have been derived.

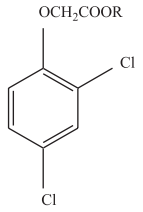
There is one expert system available for the prediction of K_{oc}, namely PCKOCWIN, developed by Syracuse Research Corporation (SRC) and freely downloadable from the E.P.A. website (www.epa.gov/oppt/exposure/docs/episuitedl.htm).

2.7 QSARs for biodegradability

Environmental risk is a function of both the intrinsic hazard posed by a chemical, and the exposure to it to which organisms (including humans) are subjected. One of the main factors influencing exposure is the length of time

that a chemical persists in the environment before being degraded. Such biodegradation is brought about very largely by bacterial attack, so that testing for biodegradation usually involves the use of a sample of soil or sewage sludge. There are a number of different types of biodegradability test, and it is important to note that biodegradability data are comparable only if determined under the same protocol. Even then, it is recognised that there is considerable error on biodegradability data. Furthermore, since different classes of chemicals are likely to have different mechanisms of biodegradation, it would be expected that the biodegradability of groups of diverse chemicals would be difficult to model using QSAR techniques; this has proved, indeed, to be the case.

A number of QSARs have been published for biodegradability of congeneric series of chemicals. Paris *et al*.[38] found that biodegradation rate constants of a small series of 2,4-dichlorophenoxyacetic acid esters (**7**) correlated well with hydrophobicity:



7

$$\log k = 0.799 \log P - 11.64 \quad (14)$$
$$n = 6,\ r^2 = 0.944,\ s \text{ not given}$$

Other studies have, however, demonstrated correlations with electronic[39] and steric[40] properties for other series of chemicals. This probably reflects different mechanisms of biodegradation of different classes of chemicals.

Diverse data-sets, however, are not very amenable to Hansch-type (*i.e.* multiple linear regression) QSAR analysis, undoubtedly because of different mechanisms of action being involved. Desai *et al*.[41] used a group contribution method to predict biodegradation rate constants of diverse chemicals, and obtained a mean error of 11.1%. A recent analysis,[42] using electrotopological state indices, found $r^2 = 0.76$ for a training set of 176 diverse organic chemicals.

The classification approach has therefore been adopted, whereby chemicals are classified as readily or non-readily biodegradable according to pre-defined criteria. For example, Dearden and Cronin[43] used discriminant analysis to model a data-set of 222 aromatic compounds; using three descriptors, they found 73.1% correct predictions for non-ready biodegradability and 88.7% correct predictions for ready biodegradability. Loonen *et al*.[44] used partial least squares discriminant analysis based on sub-structural features to obtain 84% correct predictions for ready and 86% correct predictions for non-ready biodegradability, with a large data-set of 894 compounds. Raymond *et al*.[45] have recently reviewed the QSAR prediction of biodegradability.

There are several expert systems available for the prediction of biodegradability. The SRC software BIOWIN is freely downloadable from the E.P.A. website (www.epa.gov/oppt/exposure/docs/episuitedl.htm). META (www.multicase.com) is part of the MULTICASE suite of software. It is essentially a metabolite prediction system, but has been applied to biodegradation with good results.[46] METEOR, developed by Lhasa Limited (www.chem.leeds.ac.uk/LUK) is also a metabolite prediction system; it has not yet been applied to biodegradability prediction, but there is no reason why it could not be so used.

2.8 Physicochemical property calculation

A number of physicochemical properties are of environmental importance, as the QSARs given above illustrate. Partition coefficient (P) is undoubtedly the most important of these, and there are numerous software packages available for the calculation of log P. Of these, among the best are: Interactive Analysis (www.logp.com), which allows free on-line calculation ; Biobyte (www.biobyte.com); SPARC (http://ibmlc2.chem.uga.edu/sparc), which allows free on-line calculation; and the SRC software KOWWIN, which is freely downloadable from the E.P.A. website (www.epa.gov/oppt/exposure/docs/episuitedl.htm).

Aqueous solubility is another important environmental property. The Interactive Analysis and SPARC websites allow free on-line calculation of solubility, and WSKOWWIN from SRC is freely downloadable from the E.P.A. website.

Boiling point calculation is available from SPARC (freely on-line), MPBPVP from SRC (www.epa.gov/oppt/exposure/docs/episuitedl.htm), ACD (www.acdlabs.com) and ProPred (www.capec.kt.dtu.dk/main/software/propred/propred.html). Vapour pressure calculation is available from SPARC (freely on-line) and ACD. Melting point calculation is available from MPBPVP and ProPred. Dearden[47] has recently reviewed the QSAR prediction of melting point, boiling point and vapour pressure.

The air-water partition coefficient (Henry's law constant) has important implications for the distribution

of chemicals in the environment, and numerous attempts have been made to predict it; Dearden and Schüürmann[48] have recently reviewed the QSAR prediction of Henry's law constant. The SRC software HENRYWIN is freely downloadable from the E.P.A. website.

3. Conclusions

More than 100,000 chemicals are released into the environment, and little is known about the toxicity of most of them. It would be impossibly expensive and time-consuming to test all such chemicals for toxicity. However, regulatory agencies are beginning to accept toxicities predicted by QSAR. This paper has shown that the use of QSAR for the prediction of environmental toxicity is well-established, although there is still a shortage of good quality toxicity data for the development of QSARs. Environmental fate (bioconcentration, soil sorption and biodegradation) can also be predicted by QSAR, as can physico-chemical properties of environmental relevance.

References

1. Web sites: ecb.jrc.it; www.epa.gov/opptintr/chemrtk/volchall.htm.
2. Commission of the European Communities; *White Paper on a Strategy for a Future Chemicals Policy*, Brussels, Belgium, 2001. http://europa.eu.int/comm/environment/chemicals/whitepaper.htm.
3. Toussaint, M.W.; Shedd, T.R.; Vanderschalie, W.H.; Leather, G.R.; *Environ. Toxicol. Chem.* **1995**, *14*, 907.
4. Dearden, J.C. In *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W.; Devillers, J., eds.; Kluwer Academic Publishers: Dordrecht, Netherlands, 1990, p. 25.
5. Kier, L.B.; Hall, L.H.; *Molecular Connectivity in Chemistry and Drug Research*, Academic Press: New York, 1975; Kier, L.B.; Hall, L.H.; *Molecular Connectivity in Structure-Activity Analysis*, John Wiley & Sons Inc.: New York, 1986.
6. Kier, L.B.; Hall, L.H.; *Molecular Structure Description: the Electrotopological State*, Academic Press: San Diego, CA, USA, 1999.
7. Hansch, C.; Maloney, P.P.; Fujita, T.; Muir, R.M.; *Nature* **1962**, *194*, 178.
8. Hansch, C; Leo, A.; *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*, American Chemical Society: Washington DC, USA, 1995.
9. Nendza, M.; *Structure-Activity Relationships in Environmental Sciences*, Chapman & Hall: London, 1998.
10. Cronin, M.T.D.; Dearden, J.C.; *Quant. Struct.-Act. Relat.* **1995**, *14*, 1.
11. Bradbury, S.P.; Henry, T.R.; Carlson, R.W. In *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W.; Devillers, J., eds.; Kluwer Academic Publishers: Dordrecht, Netherlands, 1990, p. 295; Bradbury, S.P.; *SAR QSAR Environ. Res.* **1994**, *2*, 89.
12. Verhaar, H.J.M.; van Leeuwen, C.J.; Hermens, J.L.M.; *Chemosphere* **1992**, *25*, 471.
13. Boxall, A.B.A.; Watts, C.D.; Dearden, J.C.; Bresnen, G.M.; Scoffin, R. In *Quantitative Structure-Activity Relationships in Environmental Sciences – VII*; Chen, F.; Schüürmann, G., eds.; SETAC Press: Pensacola, FL, U.S.A., 1997, p. 263.
14. Könemann, H.; *Toxicol.* **1981**, *19*, 209.
15. van Leeuwen, C.J.; van der Zandt, T.J.; Aldenburg, T.; Verhaar, H.J.M.; Hermens, J.L.M.; *Environ. Toxicol. Chem.* **1992,** *11*, 267.
16. Lipnick, R.L. In *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W.; Devillers, J., eds.; Kluwer Academic Publishers: Dordrecht, Netherlands, 1990, p. 281.
17. Veith, G.D.; Broderius, S.J. In *QSAR in Environmental Toxicology – II*; Kaiser, K.L.E., ed.; D. Reidel: Dordrecht, Netherlands, 1987, p. 385.
18. Cronin, M.T.D.; Schultz, T.W.; *Ecotoxicol. Environ. Safety* **1998**, *39*, 65.
19. Cronin, M.T.D.; Schultz, T.W.; *Chem. Res. Toxicol.* **2001**, *14*, 1284.
20. Mekenyan, O.G.; Veith, G.D.; *SAR QSAR Environ. Res.* **1993**, *1*, 335.
21. Hermens, J.; de Bruijn, J.; Pauly, J.; Seinen, W. In *QSAR in Environmental Toxicology – II*; Kaiser, K.L.E., ed.; D. Reidel: Dordrecht, Netherlands, 1987, p.135.
22. de Bruijn, J.; Hermens, J.; *Aquat. Toxicol.* **1993**, *17*, 257.
23. Verhaar, H.J.M.; Eriksson, L.; Sjöström, M.; Schüürmann, G.; Seinen, W.; Hermens, J.L.M.; *Quant. Struct.-Act. Relat.* **1994**, *13*, 133.
24. Devillers, J.; *SAR QSAR Environ. Res.* **2001**, *11*, 397.
25. Kaiser, K.L.E.; Niculescu, S.P.; Schultz, T.W.; *SAR QSAR Environ. Res.* **2002**, *13*, 57.
26. Dearden, J.C.; Barratt, M.D.; Benigni, R.; Bristol, D.W.; Combes, R.D.; Cronin, M.T.D.; Judson, P.N.; Payne, M.P.; Richard, A.M.; Tichý, M.; Worth, A.P.; Yourick, J.J.; *ATLA* **1997**, *25*, 223.
27. Greene, N.; *Adv. Drug Delivery Rev.* **2002**, *54*, 417.
28. Mackay, D.; *Environ. Sci. Technol.* **1982**, *16*, 274.
29. Geyer, H.; Politzki, G; Freitag, D.; *Chemosphere* **1984**, *13*, 269.
30. Dearden J.C. In *Comprehensive Medicinal Chemistry: Volume 4 – Quantitative Drug Design*; Ramsden, C.A., ed.; Pergamon Press: Oxford, UK, 1990, p. 375.
31. Deneer, J.W.; Sinnige, T.L.; Seinen, W.; Hermens, J.L.M.; *Aquat. Toxicol.* **1987**, *10*, 115.

32. Dimitrov, S.D.; Mekenyan, O.G.; Walker, J.D.; *SAR QSAR Environ. Res.* **2002**, *13*, 177.
33. Sabljić, A; *Z. Gesamte Hyg. Grenzgeb.* **1987**, *33*, 493.
34. Dearden, J.C.; Bradburne, S.J.A.; Cronin, M.T.D.; Solanki, P. In *QSAR 88*; Turner, J.E.; England, M.W.; Schultz, T.W.; Kwaak, N.J., eds.; U.S. Dept. of Energy: Oakridge, TN, USA, 1988, p. 43.
35. Briggs, G.G.; *J. Agric. Food Chem.* **1981**, *29*, 1050.
36. Hodson, J.; Williams, N.A.; *Chemosphere* **1988**, *17*, 67.
37. Sabljić, A.; *J. Agric. Food Chem.* **1984**, *32*, 243.
38. Paris, D.F.; Wolfe, N.L.; Steen, W.C.; *Appl. Environ. Microbiol.* **1984**, *47*, 7.
39. Peijnenburg, W.J.G.M.; *Pure Appl. Chem.* **1994**, *66*, 1931.
40. Paris, D.F.; Wolfe, N.L.; Steen, W.C.; *Appl. Environ. Microbiol.* **1983**, *45*, 1153.
41. Desai, S.M.; Govind, R.; Tabak, H.H.; *Environ. Toxicol. Chem.* **1990**, *9*, 473.
42. Huuskonen, J.; *Environ. Toxicol. Chem.* **2001**, *20*, 2152.
43. Dearden, J.C.; Cronin, M.T.D. In *Biodegradability Prediction*; Peijnenburg, W.J.G.M., Damborský, J, eds.; Kluwer Academic Publishers: Dordrecht: Netherlands, 1996, p. 93.
44. Loonen, H.; Lindgren, F.; Hansen, B.; Karcher, W.; Niemelä, J.; Hiromatsu, K.; Takatsuki, M.; Peijnenburg, W.; Rorije, E.; Struijś, J.; *Environ. Toxicol. Chem.* **1999**, *18*, 1763.
45. Raymond, J.W.; Rogers, T.N.; Shonnard, D.R.; Kline, A.A.; *J. Hazard. Mat.* **2001**, *B84*, 189.
46. Klopman, G.; Tu, M.H.; *Environ. Toxicol. Chem.* **1997**, *16*, 1829.
47. Dearden, J.C.; *Environ. Toxicol. Chem.*, in press.
48. Dearden, J.C.; Schüürmann, G.; *Environ. Toxicol. Chem.*, in press.

Received: June 6, 2002
Published on the web: November 18, 2002