INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

APPLIED CHEMISTRY DIVISION
COMMISSION ON WATER CHEMISTRY*

STRUCTURE–ACTIVITY RELATIONSHIPS FOR BIODEGRADATION: A CRITICAL REVIEW

(Technical Report)

Prepared for publication by

W. J. G. M. PEIJNENBURG

National Institute of Public Health & Environmental Protection, Laboratory for Ecotoxicology, P.O. Box 1, 3720 BA Bilthoven, Netherlands

Structure–activity relationships for biodegradation: A critical review (Technical Report)

Abstract - Given the general lack of experimental data, at the moment an increasing number of models are being developed that may be used to predict the environmental fate of organic chemicals, based solely on their chemical structure or physico-chemical properties. For most organic chemicals, biodegradation is the dominant environmental removal process. With the aim of assessing the limitations of Quantitative Structure Biodegradation Relationships (QSBRs) presently available, this contribution reviews the approaches used to derive QSBRs and on the results obtained. First an introduction is given on the general principles that form the fundamentals of most QSBRs developed so far, including a summary of the methods and the descriptors most commonly used. Subsequently a number of QSBRs reported in literature and their limitations are discussed. Finally, some aspects of the integrated approach that needs to be followed in order to be able to comfortably predict "real world" biodegradation rates in soil and water, is presented.

INTRODUCTION

The identification of the environmental risks of the large numbers of chemicals that are being produced and used nowadays is greatly hindered because the necessary data for emissions, fate and possible effects in the environment are lacking for the majority of these chemicals. Therefore to an increasing extent, use is made of models that merely require the input of the molecular structure (or properties derived from this structure) as the most important parameter (Quantitative Structure Activity Relationships, QSARs). At the moment an increasing number of QSAR-models are being developed that may be used to predict the environmental fate of organic chemicals. Until now several of the most relevant processes determining the environmental fate of organic chemicals have been studied in detail and various models and process descriptions are available. Generally speaking, most information is available on transport processes and QSARs for distribution processes are relatively well developed. Many of them have been incorporated into the QSAR/DATABASE [1].

In contrast to this, QSARs for environmentally relevant transformation processes are relatively underdeveloped. Contributions from abiotic and biotic degradation processes will lead to the attenuation of toxicant concentrations in the environment. The extent of these processes may vary, depending on the structural characteristics of the chemical ánd the environmental conditions. Therefore, persistence data are often difficult to compare and extrapolation of reaction rates and reaction products observed under laboratory conditions to environmental conditions appears to be only possible to a limited extent. This is due to a variety of causes; from the extreme difference in degradative mechanisms to the very specific environmental condition of each phenomenon. Nevertheless, Karickhoff et al. [2] developed a prototype computer system for the prediction of chemical reactivity strictly from molecular structure for a broad range of molecular structures.

Biodegradation is considered to be the dominant degradative route for various organic chemicals. It is important to recognize the ambiguity that may be engendered by the term 'biotic' [3]. There is a spectrum of phenomena between those that are strictly biological and those that involve only non-biological reactants. This can lead to semantic confusion. For the purpose of this contribution, the most restricted

and literal definition of biotic will be adopted: All processes directly involving the participation of metabolically active microorganisms.

In general two types of biodegradation processes may be distinguished:

1. Primary biodegradation; any biologically induced structural transformation in the parent compound that changes its molecular integrity.
2. Ultimate biodegradation or mineralization; biologically mediated conversion of an organic compound to inorganic compounds and products associated with normal metabolic processes.

Despite the important contribution of biologically mediated removal processes, generally applicable Quantitative Structure Biodegradation Relationships (QSBRs) or qualitative Structure Biodegradation Relationships (SBRs, in which the biodegradability of a compound is expressed semi-quantitatively; biodegradable or non-biodegradable) seem to be lacking to a large extent.

In this contribution a review is given on the approaches used to derive QSBRs and on the results obtained with the aim of assessing the limitations of the QSBR models presently available. However, no attempt has been made to cover the whole field exhaustively, and the presentation is therefore selective rather than comprehensive. First an introduction is given on the general principles that form the fundamentals of most QSARs developed so far, including a summary of descriptors used in 'transformation' QSARs. Also some new developments in the field of QSARs for transformation processes and some applications will be shown. Finally the major limitations of the models developed so far, will be discussed.

QSARs/QSBRs

General, methodology

As a consequence of the lack of experimental data for kinetic parameters for environmental fate processes as well as acute and chronic effects, attempts have been made to estimate the fate and effects of pollutants, in order to help assess their possible environmental hazards. Among these estimation methods the study of QSARs and of physical property-activity relationships (PARs) has attracted increasing interest. Environmental and toxicological properties are calculated from electronic, atomic, fragmental or molecular properties of molecules (molecular descriptors) or from macroscopic physicochemical properties. The basis of such calculations generally is a QSAR or PAR equation derived from environmental or toxicological properties determined for a relatively small set of reference compounds under suitable experimental conditions. The role of QSARs and PARs in environmental contaminant studies is:

1- To provide methods for reliably estimating the potential hazards of contaminants in a relatively simple way, thereby reducing the number of expensive and time-consuming experiments needed to assess those environmental hazards.
2- To provide guidelines for the classification of chemicals and to identify outliers.
3- To help understand the mechanisms of (re)action.

In order to derive reliable QSAR models, a number of essential conditions need to be fulfilled, namely [4]:

- Similarity in structure and mechanism of action of the compounds used as a training set.
- Characterization of the structure with consistent, relevant descriptors.
- A representative training set.
- Data analysis with statistically sound methods.
- Homogeneously measured test data.
- Internal and external validation of the model obtained.

In general three broad classes of data analytical methods can be distinguished within QSAR [5]:

1 - For homologous substances, SARs have been developed by univariate methods (multiple regression analysis).
2 - Multivariate processes (like discriminant analysis, pattern recognition techniques, principal component analyses and factor analyses) are employed in order to derive structure based models, which are intended for application to various chemical classes.
3 - Diverse substructures associated with degradable and non-degradable compounds are used to qualitatively approximate potential degradability of substances (cluster and ranking analysis).

Multiple regression is the most common method applied in QSAR studies within environmental chemistry and toxicology. The basis of the method as originally developed by Hansch and Fujita [6] is the postulated relationship between a certain property (CP) and structure descriptors that influence it (for example, P, Q and R), as pictured in the following equation, in which both k, a, b and c are constants:

$$[CP] = k \times [P]^a \times [Q]^b \times [R]^c \qquad (1)$$

Or, after log-transformation:

$$\log [CP] = a \times \log [P] + b \times \log [Q] + c \times \log [R] + \log [k] \qquad (2)$$

This equation can be solved by multiple linear regression. The approach assumes that the effect of a substituent at a specific position is constant and independent of the effect of substituents at other positions. Additivity will fail if there are interactions between certain of the substituents. A rationalization of equation 1 may well be given in terms of linear free energy relationships.

The general procedure for deriving a QSAR using multiple regression techniques is as follows; first, a group of related compounds (training set) is selected. Then, the property of these compounds that has to be described, is measured (or these data are taken from the literature) and one or more physico-chemical parameters of which it is assumed that they might play a role are either calculated or measured. Then the (logarithms of the) values obtained are subjected to linear regression, in order to calculate the best-fitting values of the constants. Finally, the model thus obtained needs to be validated with the data set used (internal validation) as well as with additional data sets (external validation). As noted by Eriksson et al. [7], especially the selection of the training set and the validation of the predictive ability of the QSAR model are largely neglected.

In figure 1 the general outline on the procedures used for deriving structure activity relationships is given.
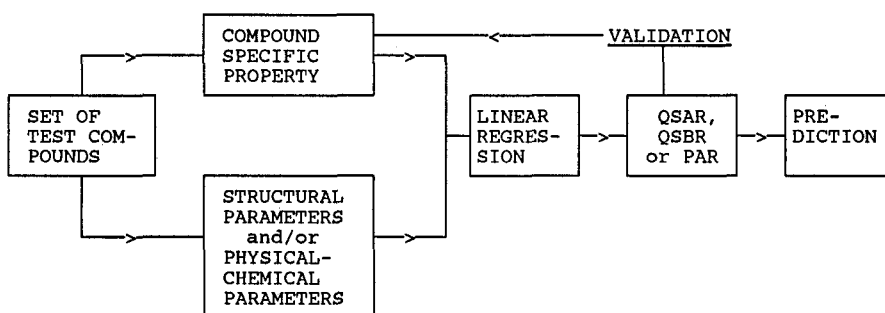

FIGURE 1. General outline on the procedures used for deriving structure activity relationships.

It should be noted that no fundamental theory exists for the formulation of QSAR-compatible classes. It may be anticipated that the structural range covered by a QSAR will be largely dependent on the specificity of the mechanism underlying the endpoint of interest. Thus, more than one organic class may be included in the same model in the case of a poorly specific transformation process; vice versa, limited ranges of structural variability within a class may be compatible with the modeling of highly specific processes. In all cases, the assessment of the actual range of a QSAR will result from data analysis and model validation [8].

Descriptors
The derivation and application of SARs require descriptors for molecules and molecule fragments, for example: physico-chemical, geometric, electronic and topological parameters. A detailed classification of molecular descriptors is given by Govers and de Voogt [9,10]. Molecular descriptors are properties, or characteristics, inherent to a molecule or to its constituting parts, which together represent the entire molecule. According to Govers and de Voogt, classification can be based on the method by which the values of the molecular descriptors are obtained, or on the specific system (atom, fragment, molecule, intermolecular complex, macroscopic system) to which they are directly related.

An alternative, less rigid method implies a classification along two lines:

- **the specific property** the descriptor is referring to:
    - A. the structure or
    - B. the energetics/interactions of a system
- **the system level** which is used to understand it:
    1. electronic
    2. atomic/fragmental/molecular
    3. intermolecular
    4. macroscopic physicochemical or biological

A. structure descriptors

The structure of a molecular system comprises the way it is constructed or constituted from smaller units. It specifies the exact relative positions of these units and the exact dimensions (internal and external geometry) of the system, or uses topological indices to describe the constituting units, their relative positions and dimensions.

1 - Electronic structure descriptors represent the number of electrons and describe the way these are distributed in the molecule and its atoms; they can be calculated directly by topology or by quantum mechanics. Structure-related molecular properties also belong to this category. Many of these properties can also be obtained from macroscopic properties determined experimentally. Some well-known electronic structure descriptors include: Valence connectivity index, electronic charges on atoms, electric moments (for instance, dipole moment), polarizability.

2 - Atomic/fragmental/molecular structure descriptors represent the number of atoms and atom groups and/or describe the way in which molecules are constructed from these. They may differ with respect in the degree to which the constituting atoms and their bonds are specified and quantified. Quantum mechanical, semi-empirical (atom-atom potential), mathematical methods and experimental measurements are used to obtain these descriptors.

Two sub-classes can be distinguished: geometric and topological descriptors. Geometric descriptors are calculated using the exact position of atomic centers relative to each other (non-steric descriptors referring to the internal structure) and/or rigid atomic radii (steric descriptors referring to the external structure). Symmetry descriptors can also be used to this end. They describe the operators by which a molecule can be superposed on itself. The geometry of molecules does not change during these operations, but need not be specified either. Geometric structure descriptors include the length:width ratio, distance index, van der Waals Volume and steric overlap volume [10].

Topological descriptors describe those structural properties of molecules, which do not change when a molecule is deformed without cleavage or superposition of bonds. Topological descriptors include the number of atoms, or of atom groups, and the connectivity indices. Connectivity indices specify the topology of a molecule. Each atom of a molecule is assigned a number (δ) corresponding to the number of adjacent non-hydrogen atoms. The 1st order connectivity index (^1X) is obtained via summation of the products $(\delta_i \cdot \delta_j)^{-1/2}$ for all pairs of adjacent atoms i and j. The 2nd order connectivity index (^2X) is obtained from all δ products of three neighboring atoms [11].

3 - Intermolecular structure descriptors specify the relative positions and orientations of molecules to each other, for instance, the way in which biologically active compounds fit into a receptor cavity or the way molecules are linked to each other. Quantum chemical, geometric and semi-empirical (atom-atom potentials) methods and experimental measurements are used to obtain these descriptors.

4 - Macroscopic structure descriptors specify the structure of macroscopic phases. They can be used to calculate genuine (structure) descriptors such as mentioned above and below, and to predict physicochemical properties (PAR studies). Usually they are derived using experimental methods. A well known macroscopic structure descriptors is: molar volume/molar density.

B. (Free) energy descriptors

Free energy descriptors specify interactions between the elements of a certain system or between the system and its environment [10]. At a molecular or submolecular level, when dealing with small numbers of particles, only (quantum-mechanical) energy is relevant. At the macroscopic level, or when dealing with large numbers of particles, (statistical) thermodynamic (free) energy or enthalpy and entropy are relevant.

Entropy describes the degree of order of the system and, like for instance molar volume, can be classified as a (macroscopic) structure descriptor itself. Moreover, some thermodynamic expressions of entropy closely resemble those from information theory.

Electronic (free) energy descriptors specify the electronic energy of atoms, bonds and molecules including the attraction and repulsion of electrons by atoms and molecules. They are calculated by quantum-mechanical methods or from properties determined experimentally. Examples of electronic energy descriptors are: electronegativity, ionization potential, electron affinity, Hammett electronic constant, reduction potential, and Hückel π-binding energy.

Atomic/molecular (free) energy descriptors (like bond strength, intermolecular forces, intra-molecular atom-atom potentials), reflect the strength of atomic bonds and the attractive and repulsive interactions between atoms in molecules. They are related to the chemical stability and reactivity of molecules. These properties can be calculated by quantum-mechanical and semi-empirical (atom-atom potential) methods or derived from macroscopic properties which have been determined experimentally.

Intermolecular (free) energy descriptors (like Taft constants, potential energy field parameters, and reversed phase HPLC retention index), describe the van der Waals (electrostatic, inductive and dispersive) interaction, intermolecular H-bond energy, intermolecular charge-transfer energy and/or repulsive energies. These properties can be calculated by quantum-mechanical and semi-empirical (atom-atom potentials, Lippingcot-Schröder potentials) methods and from experimentally determined macroscopic properties. For instance, the Hammett electronic constant and the Taft steric constant are derived from (partly) intra-molecular processes (dissociation of benzoic acid and ester hydrolysis, respectively).

Macroscopic (free) energy descriptors (like n-octanol/water partition coefficients, chromatographic retention parameters, acid/base dissociation constants (pK_a) and ester hydrolysis constants), are used for the calculation of molecular descriptors as described above or applied in PAR studies.

The following descriptors are used most frequently for deriving QSARs for environmental transformation processes:
- The summation of the Hammett-substituent constants (σ) [12]. Hammett-substituent constants specify the electron-attracting or electron-repelling effect of substituents. σ can be regarded as an approximate measures of the relative electron density at the center of reaction. For polysubstituted compounds it may be assumed that the individual effects are additive when strong interactions do not prevail.
- The summation of the inductive constants of the additional substituents, σ_I [12]. σ_I describes the intrinsic tendency of a substituent to withdraw electrons, σ_I being positive for electronegative groups as compared to hydrogen. The effect weakens steadily with increasing distance from the reaction center. It should be noted that σ_I-values to a certain extent are correlated to the corresponding Hammett σ constants.
- In case of aliphatic compounds; the Taft sigma constant of the substituents (σ^*). This parameter too is used to describe electronic effects.
- The Taft steric constant, E_s [13]. This being the classical descriptor for the steric effects of substituents on chemical and biological processes.
- In case hydrophobicity or transport to an active site is involved, the n-octanol-water partition coefficient (K_{ow}) is often used. A detailed discussion on methods used to measure or calculate K_{ow} is given by Hermens [5].
- The pK_a-value of corresponding benzoic acids; this under the assumption that the influence of substituents on reactivity is comparable to their influence on pK_a-values.
- Parameters calculated by quantum chemical methods, such as the extended Hückel theory [14] or semi-empirical methods (MNDO [15]; AM1 [16]; CNDO/2 [17]). These include ionization potentials (IP), dipole moments, charge densities, HOMO and LUMO energies (Highest Occupied Molecular Orbital, Lowest Unoccupied Molecular Orbital) and atomic charges [18]. These parameters are dependent on the conformation of the molecules. All calculation processes are based on the assumption that the most probable conformation is the conformation involving the minimum amount of energy. However, depending on the method employed, varying results are obtained.
- Other macroscopic descriptors, including abiotic reaction rate constants.

BIODEGRADATION AND RELATIONSHIPS FOR BIODEGRADATION PROCESSES

Biodegradation, general
Metabolism by microorganisms is one of the most important processes determining the fate of chemicals in the environment. In some cases, microbial populations use synthetic organic chemicals as sources of dietary energy. When this is so, classical microbial techniques, including Michaelis-Menten-Monod equations for growth kinetics of bacterial populations, can be used in chemical fate modelling. The results of laboratory investigations in which the synthetic compound is used as a sole source of carbon and dietary energy can be difficult to extrapolate to the field situation, however. The population dynamics, viability, and metabolic status of specialized subpopulations of microbial degraders cannot currently be predicted with much confidence. In addition, in many cases synthetic compounds are degraded by multispecies microbial assemblages under natural conditions, despite the fact that no individual species capable of using the compound as a sole carbon source can be isolated. Rates of biodegradation are a

function of the rates of a series of processes. Correlations between biodegradation rates and structure will be facilitated if one of these processes is rate determining. In general, these processes will be of two types. Biodegradation rates may be determined by rates of uptake and transport (like for example rates of uptake by microbial cells or rates of transport within the cell to the relevant enzymes), by the rate of binding to the active site of an enzyme, and/or by the rate at which they undergo enzymatic transformation [19].

In the absence of a specific uptake mechanism, synthetic organic chemicals are probably transported into bacterial cells by passive diffusion through the lipid membrane. The permeability coefficient for diffusion is proportional to the lipid-water partition coefficient if the concentrations just within the membrane are related to those in the adjacent aqueous phases by this partition coefficient [20]. Similar relationships are found between permeability coefficients and other parameters describing hydrophobicity or lipophobicity, such as the *n*-octanol-water partition coefficient [20,21]. Therefore, if biodegradation rates are limited by the rates of uptake by diffusion through the cell envelope, one would expect to find correlations between biodegradation rate constants and macroscopic hydrophobic parameters such as the *n*-octanol-water partition coefficients.

The enzyme-catalyzed transformation of a compound is preceded by its binding to the active site of the enzyme by the formation of hydrogen or covalent bonds. The strength of this interaction is influenced by the electronic structure of the compound, as well as by steric effects on the fit of the compound on the active site. The reactivity of the bound compound will also be determined by its electronic and steric properties. Consequently, if enzyme binding or transformation rates determine biodegradation rates, correlations would be expected between biodegradation rate constants and molecular factors influencing the binding or the reactivity of the chemicals, i.e., with steric or electronic parameters. However, it is not always possible to distinguish between these possibilities. For example, steric factors will not only influence the enzyme binding and reactivity, but also the hydrophobicity of a compound.

Relationships for biodegradation

Extended reviews on structure-activity relationships between biodegradation rates of organic compounds and chemical structure parameters are (amongst others) given by Degner et al. [18] and by Parsons and Govers [19]. In general, two approaches may be distinguished:

1 - A mechanistic approach in which it is attempted to identify the rate-limiting step within the overall process [19]. In this approach often simple or multiple regression analysis is used to develop the QSBR models.
2 - A more pragmatic approach, sometimes referred to as a black-box approach, in which large datasets containing results of various test procedures are used as a training-set [22,23]. Chemicals tested are subdivided into specific chemical classes and subsequently SAR models for estimating the microbial degradability of homologous substances are developed. Often use is made of elaborate statistical methods to derive the SBRs, like neural networks and multivariate data-analytical methods. The outcome of these methods often is a boolean classification (yes/no biodegradable) or a ranking of biodegradation potential; SBR models.

In both cases the biodegradation parameter used as dependent variable may vary, examples of which include: percentage of decomposition after a given time period, percentage of theoretical oxygen demand, rate of biodegradation, and five-day biological oxygen demand [22].

In the contribution of Degner et al. [18] it was concluded that, though numerous publications have been published on this subject, structure activity relationships to predict biodegradation are only rarely suited to be used by the competent authorities in the course of chemical assessment procedures. This may well be due to the fact that most of the structure activity relationships are only valid for small groups of molecules of distinct chemical classes and therefore some expertise is needed to select an appropriate equation. A validation study of SARs for estimating biodegradation revealed that only few SAR models can be recommended for predictive purposes under environmental conditions [18]. A considerable number of SARs (64) was derived with compounds that are classified by the MITI test [24]. Most models, however, were either incapable of discriminating degradability, or could not be validated. Therefore Degner et al. developed further models using several weighted substructure indicators, assuming that indicator variables for functional groups are to reflect enhancing/retarding effects on transformation rates. However, due to the high degree of uncertainty involved in the results obtained, application of these models to estimate biodegradation cannot be recommended [18]. Validation yielded satisfactory agree-

ment between predicted and experimental degradation data only for 12 out of a total of 78 models. These models are comprised in decision-nets for acyclic, monocyclic aromatic and aliphatic cyclic compounds. The degradability of compounds belonging to other chemical classes can be predicted only at a low level of reliability at present.

1 - QSBR-models

Parsons and Govers [19] give a systematic review of QSARs for biodegradation kinetics; three types of descriptors are distinguished:
 a - Macroscopic parameters.
 b - Molecular electronic parameters.
 c - Molecular steric parameters.
Which descriptor gives the best correlation depends on the mechanism of the biodegradation rate-determining step. Normally, since rates of biodegradation depend on both chemical characteristics and on the activity of the microbial population capable of degrading a chemical, second-order rate constants are required as biodegradation rate parameters in QSBRs. However, there is a limited number of such data available.

a - Relationships between biodegradation rates and macroscopic parameters.

The most commonly used structural descriptors are macroscopic parameters that describe hydrophobicity, in particular n-octanol-water partition coefficients [25] and reverse-phase HPLC retention times [26]. Other macroscopic descriptors include abiotic reaction rate constants. Reversed-phase HPLC retention correlates quite well with partition between water and organic phases such as n-octanol and is sometimes used as an alternative hydrophobicity parameter. A decrease in biodegradation rate for the most hydrophobic compounds was found, which was interpreted as indicating that uptake of these compounds by diffusion is the process determining their biodegradation rate. However, one would in general expect increasing hydrophobicity to lead to increasing uptake rates, and thus increasing biodegradation rates if uptake is rate-determining.

Banerjee et al. [27] measured the relative second-order biodegradation rate constants of chlorinated phenols, anisoles, and resorcinols in pure cultures and in samples of natural waters. The data showed a trend of decreasing biodegradation rate constants with increasing n-octanol-water partition coefficients. This observation was explained in terms of a kinetic model for biodegradation where uptake of compounds by bacteria is the rate-determining process. The model includes terms for sorption of compounds to the cell wall and uptake by diffusion through both hydrophilic pores and lipid layers. It was concluded that rate constants for diffusion through the lipid layers are inversely proportional to n-octanol-water partition coefficients. Also, decreasing reactivity of the enzyme-chlorophenol complexes with increasing log K_{ow} may have been responsible for the observed trend in biodegradation rates [28].
In general rates of diffusion of organic compounds through lipid membranes increase with increasing hydrophobicity and decrease with increasing size.

b - Relationships between biodegradation rates and molecular electronic parameters.

Electronic parameters used for deriving QSBRs include:
 - The substituent's Hammett constants. A correlation between chemical oxygen demand and the substituent's Hammett constants had a negative slope, indicating rate-limiting electrophilic attack [29]), whereas also correlations between the Michaelis constant and Hammett constants are reported for dioxygenation of substituted catechols [30].
 - The difference in modulus of atomic charge across one specific bond ($\Delta[\sigma]_{X-Y}$). This descriptor was used by Dearden and Nicholson as descriptor for the 5-day biological oxygen demand (BOD) of various classes of chemicals [31].

c - Relationships between biodegradation rates and molecular steric parameters.

Correlations of second-order biodegradation rate constants of para-substituted phenols in pure bacterial suspensions with different physicochemical parameters were compared by Paris et al. [32]. The physicochemical parameters included the pK_a, the Hammett substituent parameter, the Taft's steric parameter, log K_{ow}, and the substituents van der Waals radii. The best results were obtained with the van der Waals radius. These results suggest that biodegradation rates of the compounds studied are controlled by steric properties of the substituents, which may affect the binding of these compounds to enzymes.

Biodegradability data for different classes of compounds were correlated with molecular connectivity indices by Boethling [33]. Although this approach yielded good correlations for a wide range of compounds, the fact that for different structural classes different rate parameters and different order connectivity indices were used to obtain the best correlations limits their predictive utility, and also makes it difficult to draw general conclusions concerning the influence of molecular structure on biodegradation rates. Furthermore, molecular connectivity indices themselves often correlate well with parameters expressing hydrophobicity, such as *n*-octanol-water partition coefficients or soil sorption coefficients.

2 - SBR-models

The ranking of substances in their relative order of biodegradability is more valuable than boolean classification to establish structure-biodegradability relationships. Such studies imply an evaluation in identical experimental conditions and generally concern substances derived from the same chemical structure and bearing different substituents. According to Kuenemann et al. [22], one advantage of SBR studies over QSBR studies is that more chemicals can be introduced in the training set because less stringent requirements are placed on the data (e.g. biodegradable or non-biodegradable, according to different end-points). Niemi et al. [34] used discriminant analysis to assess the structural features associated with the biodegradation of chemicals tested with the standard 5-d BOD procedure. The multivariate analysis was based on 54 molecular connectivity indices and a limited number of physicochemical properties. With the best iteration, 85 % of the degradable chemicals and 94 % of the persistent chemicals were correctly predicted. Based on these results and based on results presented in literature, a series of features associated with degradable and persistent chemicals were identified. This approach allowed correct prediction of 91 % of the degradable and 96 % of the persistent compounds. In order to minimize the number of false negatives and false positives Kuenemann et al. propose to test other molecular descriptors and new statistical methods [22].

Howard et al. [23] used a file of evaluated biodegradation data [35] to develop a model for predicting aerobic biodegradability from chemical substructures. To this end, chemicals were initially divided into three groups:

1 - Chemicals that degrade rapidly under most environmental conditions without requiring acclimation.
2 - Chemicals that are biodegradable, but only after an acclimation period.
3 - Chemicals that degrade slowly or not at all.

With the chemicals from the first and the last group, two models were developed for classifying chemicals as rapidly biodegradable or not rapidly biodegradable; one based on linear and one on nonlinear regression. Both models predict the biodegradation categories correctly for 90 % of the chemicals of the training set and an independent test set. However, the number of correct predictions for chemicals that biodegrade rapidly exceeds the number of correct predictions for not rapidly degrading compounds (95 versus 80 %).

DISCUSSION

From the foregoing it may be concluded that, though various encouraging results have been obtained, the development of QSBRs is still in its initial stages. There is no general relationship between biodegradability and chemical structure. In general it is difficult to identify the rate-determining step in the degradation. Also the fact that quantitative relationships between biodegradation rates and chemical structure are in general specific to limited numbers of structurally related compound does limit their general predictive utility. In fact, there are no reports of the use of such relationships to successfully predict biodegradation rates of compounds of unknown biodegradability. There is also little information on the use of QSBRs based on biodegradability data from one system to predict biodegradation rates in other systems. More reliable data on biodegradation rates and a better understanding of the mechanism and the rate-determining steps of biodegradation are required if (Q)SBRs are to be more than a "black box" tool.

Despite the fact that biodegradability studies are the subject of a large amount of research, at present the possibilities of applying QSBRs to practical problems are limited. Above all an integrated approach needs to be followed in order to be able to reliably predict "real world" biodegradation rates in soil and water. To this end, amongst others, the following major limitations of the present QSBR models need to be addressed:

1 - Development of QSBRs is hampered by a lack of well documented data measured under uniform conditions. Especially for screening tests the results are highly dependent upon the test protocol. Often important details affecting reproducibility are not reported. Howard et al. [35] developed an evaluation method for collecting and evaluating individual studies and test data that facilitates the development of conclusions on biodegradability for use in structure/biodegradation relations [35]. This was done under the assumption that as the number of consistent test results or test results for which apparent inconsistencies are resolvable increases, the greater is the likelihood that the indication of biodegradability is a property of the chemical rather than of the test system.

2 - There is a clear lack of understanding of the variability of the microenvironments in which biodegradation can be important. This means that extrapolation of results of biodegradation tests from one ecosystem to another is possible only to a limited extent. The extrapolation of results obtained under laboratory conditions to real world situations especially is hampered by a lack of knowledge of the factors influencing biodegradation processes. Thus, differences of over two orders of magnitude were found for the half-life for disappearance of methylene chloride in a variety of surface and subsurface soils [36,37].

3 - Related to this is the problem of the various endpoints that are used as dependent parameters in (Q)SBRs. Often, the endpoints selected cannot be extrapolated to give information on biodegradation processes under realistic environmental conditions. This implies that a compound predicted to be readily degradable according to a certain parameter, is not necessarily readily degradable under realistic environmental conditions. Also additional factors such as the duration of biodegradation tests performed vary from one test system to another.

4 - Insufficient attention has been paid to the methods used for deriving robust QSBRs, broadly applicable to various groups of chemicals. As shown above, most QSBRs are applicable only within narrow compound families in which the chemical structures are fairly similar. This necessitates the use of an increasing number of molecular descriptors whilst their definitions and similarities often are obscure. In general the understanding of the fundamentals of the current QSAR techniques is insufficient to accommodate chemicals sufficiently different in structure from those upon which the methods are based; in practice 'similarity' of structure can only be generated post hoc by reference to agreement with a given QSAR. Given the various functional groups that generally are present within a molecule it often is difficult to classify a given chemical as belonging to a certain group of chemicals.

5 - The methodologies for deriving QSBRs need to be improved. This includes the selection of both the training and the validation set, the development and selection of 'better', more appropriate and broadly applicable descriptors for the transformation processes taken into consideration, the introduction of more sophisticated statistical analytical techniques and the internal and external validation (range of application and predictive capability), of the models obtained. For the validation of the models obtained it is important to note that in case of classification of chemicals as being degradable or persistent, false negatives and false positives do not have the same meaning. For ecosystems, considering a persistent chemical as being degradable could lead to problems is far worse than considering degradable compounds as persistent.

In general it may be concluded that in order to make a significant step forward in deriving structure biodegradability relationships, an _integrated_ approach needs to be followed, covering all aspects involved in deriving and applying structure biodegradability relationships.

ACKNOWLEDGEMENT

REFERENCES

1.   The QSAR SYSTEM, a structure-activity based chemical modeling and information system (data base) (1986). U.S. Environmental Protection Agency, Environmental Research Laboratory, Duluth, Minnesota.

2. S.W. Karickhoff, V.K. McDaniel, C. Melton, A.N. Vellino, D.E. Nute and L.A. Carreira, Environ. Toxicol. Chem., 10, 1405-1416 (1991).
3. D.L. Macalady, P.G. Tratnyek and T.J. Grundl, J. Contamin. Hydrology, 1, 1-28 (1986).
4. M.L. Tosato, C. Chiorboli, L. Eriksson and J. Jonsson, Sci. Total Environ., 109/110, 307-325 (1991).
5. J.L.M. Hermens, The handbook of environmental chemistry, vol. 2 part E. Reactions and processes, O. Hutzinger (ed.), pp. 111-162, Springer-Verlag, Berlin (1989).
6. C. Hansch and T. Fujita, J. Am. Chem. Soc., 86, 1616-1626 (1964).
7. L. Eriksson, J. Jonsson and R. Berglind, Environ. Toxicol. Chem., 12, 1185-1191 (1993).
8. M.L. Tosato, QSAR: rational approaches to the design of bioactive compounds, vol. 16, C. Silipo and A. Vittoria (eds.), pp. 501-504, Pharmaco-chemistry Library, Elsevier, Amsterdam, Netherlands (1991).
9. P. de Voogt, QSARs for the environmental behaviour of polynuclear (hetero)aromatic compounds, Ph. D. Thesis. Free University of Amsterdam, Amsterdam, 1990.
10. H. Govers and P. de Voogt, QSARs in environmental toxicology - III, J.E. Turner, M.W. England, T.W. Schultz and N.J. Kwaak (eds.), pp. 181-183, proceedings of the workshop "QSARs in environmental toxicology - III", Springfield, VA, May 22-26, 1988 (1988).
11. L.B. Kier and L. Hall, Molecular connectivity in chemistry and drug research,, Academic Press, New York, NY (1979).
12. C. Hansch and A.J. Leo, Substituent Constants for Correlation Analysis in Chemistry and Biology, John Wiley & Sons, New York, NY (1979).
13. R.W. Taft, Steric effects in organic chemistry, M.S. Newman (ed.), John Wiley, New York, NY (1956).
14. R. Hoffmann, J. Chem. Phys., 39, 1397-1412 (1963).
15. M.J.S. Dewar and W. Thiel, J. Am. Chem. Soc., 99, 4899-4907 (1977).
16. M.J.S. Dewar, E.G. Zoebisch, E.F. Healy and J.J.P. Stewart, J. Am. Chem. Soc., 107, 3902-3909 (1985).
17. J.A. Pople, D.P. Santry and G.A. Segal, J. Chem. Phys., 43, S129-S135 (1965).
18. P. Degner, M. Müller, M. Nendza and W. Klein, Estimating biodegradability of chemicals by computer assisted reactivity simulation; SARs on biodegradation, The Federal Environmental Agency of Germany (ed.), Berlin, Germany (1991).
19. J.R. Parsons and H.A.J. Govers, Ecotox. Environ. Safety, 19, 212-227 (1990).
20. W.D. Stein, Transport and diffusion across cell membranes, Academic Press, New York, NY (1986).
21. H. Nikaido and M. Vaara, Microbiol. Rev., 49, 1-32 (1985).
22. P. Kuenemann, P. Vasseur and J. Devillers, Practical applications of quantitative structure activity relationships (QSAR) in environmental chemistry and toxicology, W. Karcher and J. Devillers (eds.), pp. 343-370, Kluwer academic publishers, Dordrecht, Netherlands (1990).
23. P.H. Howard, R.S. Boethling, W.M. Stiteler, W.M. Leylan, A.E. Hueber, J.A. Beauman and M.E. Larosche, Environ. Toxicol. Chem., 11, 593-603 (1992).
24. OECD, Guidelines for testing chemicals, -section 3: Degradation, OECD, Paris, France (1989).
25. Y. Yonezawa and Y. Urushigawa, Chemosphere, 8, 139-142 (1979).
26. Y. Urushigawa and Y. Yonezawa, Chemosphere, 8, 695-703 (1979).
27. S. Banerjee, P.H. Howard, A.M. Rosenberg, A.E. Dombrowski, H. Sikka and D.L. Tullis, Environ. Toxicol. Chem., 3, 416-422 (1984).
28. P.L. Beltrame and P. Carniti, Environ. Toxicol. Chem., 4, 374-375 (1985).
29. P. Pitter, Acta Hydrochim. Hydrobiol., 13, 453-460 (1985).
30. E. Dorn and H.-J. Knackmuss, Biochem. J., 174, 85-94 (1978).
31. J.C. Dearden and R.M. Nicholson, Pestic. Sci., 17, 305-310 (1986).
32. D.F. Paris, N.L. Wolfe and W.C. Steen, Appl. Environ. Microbiol., 44, 153-158 (1982).
33. R.S. Boethling, Environ. Toxicol. Chem., 5, 797-806 (1986).
34. G.J. Niemi, J.D. Weith, R.R. Regal and D.D. Vaishnav, Environ. Toxicol. Chem., 6, 515-527 (1987).
35. P.H. Howard, A.E. Hueber and R.S. Boethling, Environ. Toxicol. Chem., 6, 1-10 (1987).
36. J.W. Davis and S.S. Madsend, Environ. Toxicol. Chem., 10, 463-474 (1991).
37. P.H. Howard, Environ. Toxicol. Chem., 12, 1135-1137 (1993).