QUANTITTATIVE STRUCTURE-ACTIVITY RELATIONSHIPS OF COMPARATIVE TOXICITY TO AQUATIC ORGANISMS

by

MARK TIMOTHY DAVID CRONIN

A thesis submitted in partial fulfilment of the requirements of the Council for National Academic Awards for the Degree of Doctor of Philosophy

December 1990

School of Health Sciences,
Liverpool Polytechnic,
Byrom Street,
Liverpool,
L3 3AF.

(in collaboration with)
Water Research Centre,
Henley Road,
Medmenham,
Marlow,
Bucks,
SL7 2HD.

Acknowledgements

I would like to express my sincere thanks to my supervisor Professor J.C. Dearden for his advice and inspiration throughout the project.

I am also indebted to Dr A.J. Dobbs of the Water Research Centre for his help, and financial support with the project.

I thank Mr M. Gardener, and the other members of the inorganic chemistry section at the Water Research Centre, for their help with the experimental work in the project.

I thank the various members of the Computer Services Department at Liverpool Polytechnic who have assisted me with many aspects of this project. I am grateful to Mr J. Higgins of the Mathematics Department at Liverpool Polytechnic for his comments on statistics.

I acknowledge receipt of a research studentship from the Natural Environment Research Council.

Finally, I would like to thank all of my fellow research students, especially Sue Bradburne, Nicola Gregg, and Hugh O'Neill for their friendship, and very occasionally for their helpful advice.

Summary

Quantitative Structure-Activity Relationships (QSAR) attempt statistically to relate the physico-chemical properties of a molecule to its biological activity. A QSAR analysis was performed on the toxicities of upto 75 organic chemicals to two aquatic species, Photobacterium phosphoreum (known as the Microtox test), and the fathead minnow. To model the toxicities 49 physico-chemical and structural parameters were produced including measures of hydrophobicity, molecular size and electronic effects from techniques such as computational chemistry and the use of molecular connectivity indices. These were reduced to a statistically more manageable number by cluster analysis, principal component analysis, factor analysis, and canonical correlation analysis. The de-correlated data were then used to form relationships with the toxicities. All the techniques were validated using a testing set. Some good predictions of toxicity came from regression analysis of the original de-correlated variables. Although successful in simplifying the complex data matrix, principal component analysis, factor analysis, and canonical component analysis were disappointing as predictors of toxicity. The perfomance of each of the statistical techniques is discussed.

The inter-species relationships of toxicity between four commonly utilised aquatic endpoints, fathead minnow 96 hour LC_{50}, Microtox 5 minute EC_{50}, Daphnia magna 48 hour LC_{50}, and Tetrahymena pyriformis 60 hour IG_{50}, were investigated. Good relationships was found between the fathead minnow and both T. pyriformis and D. magna toxicities indicating that these species could be used to model fish toxicity. The outliers from individual relationships were assessed in order to elucidate if any molecular features may be causing greater relative toxicity in one species as compared to another. It is concluded that in addition to the intrinsic differences between species, the greater length of the test time for any species may result in increased bioaccumulation, metabolism, and detoxification of certain chemical classes. The relationships involving fish toxicity were moderately improved by the addition of a hydrophobic parameter.

Contents

List of Figures

List of Tables

1.1. INTRODUCTION

Toxicology, or the so-called 'study of poisons', can be traced back over 3500 years. The first evidence of knowledge of the potentially harmful effects of chemicals was reported in 1500 BC by the ancient Egyptians in the Ebers Papyrus (the earliest written medical records) as references to, and recipes for, poisons. Through history since that time, the danger of chemicals introduced into our environment has been noted. Arsenic, aconite, and opium are described by the Hindu Vedas in 900BC; the ancient Greeks positively promoted the research and development of the antidotes of poisons; and the Romans first legislated against poisoning in 82BC in an attempt to curb careless dispensing. During the 'Dark' and 'Middle' Ages the study of toxins became an art. Stories of their wrongful use, for personal gain, abound (Timbrell 1989).

Since such socio-economic events as the industrial revolution in the nineteenth century, the danger to man and his environment from chemicals has further increased, and their study is of vital importance. Present day estimates suggest that there are at least 70,000 man-made and natural chemicals in everyday use. Between 500 and 1000 new chemicals are added to this list annually. They may reach our environment as pollution from a variety of sources, and if in sufficient quantities will act as poisons. Toxins are found in our food, water, air, and soil, originating as drugs, food additives, industrial outputs, agrochemicals, and even our common household products. It is the danger of man's exposure to these chemicals, their conceivable toxic or poisonous effect, that the modern-day toxicologist is concerned with. Chronic and acute exposure to chemicals can prove dangerous, so safe levels of exposure need to be determined and applied for our well being.

The understanding and regulation of a chemical's toxic effect is thus essential, and recent legalisation decrees statutory testing to obtain such information. In the United States of America under Section 5 of the Toxic Substances Control Act of 1977, the U.S. Environmental Protection Agency (EPA) is responsible for reviewing the potential hazard of new industrial chemicals to human health and to

the environment prior to production. Similarly, within the European Community, the testing, notification and classification of new chemicals are based on a set of comprehensive and detailed directives. European Directive 79/831B/EEC, for instance, dictates the obligatory notification of new chemicals.

Traditionally, obtaining accurate and meaningful information to comply with the regulations requires the extensive use of animal tests. However, the cost and time involved of individual tests is prohibitive even for new chemicals, without consideration of the vast number of commonly used existing chemicals for which no toxicological data exist. Conservative estimates by the EPA put the cost of an acute dermal toxicity test at $2500-$4500; $20000-$25000 for a 14-day inhalation study; $300000-$400000 for a two-year dietary study; and up to $1500000 for a two-year inhalation study. In addition public opinion, which is swinging away from the unnecessary use of animal experimentation and the suffering it may cause, has reinforced the need to find reasonable alternatives. Research into rapid, and cheap, alternatives has led to the investigation of techniques such as tissue culture and, of increasing interest in environmental toxicology, Quantitative Structure-Activity Relationships (QSAR), as well as the extrapolation of toxicity data from lower organisms to higher species.

The study of QSAR is based on the rationale that the effect of a chemical within a biological system (e.g. its toxic action) is solely dependent on its chemical composition. It has been successfully applied to the pharmaceutical and pesticide industries for some years to optimise drug design studies. Its main features are described below, as well as the historical aspects, and their application in modern environmental toxicology.

1.1.1. Quantitative Structure-Activity Relationships

A convenient summary of a Quantitative Structure-Activity Relationship is that it attempts statistically to relate the physico-chemical properties of a molecule to its biological activity. Obviously there are three components essential to this statement - the three components needed to create a QSAR:

i) the biological data for a set of chemicals which is to be modelled;

ii) some physico-chemical descriptors of the chemicals;

iii) a statistical technique to relate the activity to the chemical descriptors.

Each of these areas is discussed below in detail.

1.2 Biological Data Utilised in QSAR Analysis

The prediction and explanation of biological data is, of course, the crux of any QSAR analysis. Before a strong model can be formed a reliable, high quality data set of biological activities is required on which to base the relationship. Indeed it is true to say that the model can only be as good as the original biological data. Also it must be remembered that whereas chemical descriptors can, in principle, be determined or expressed with good accuracy, the qualitative and/or quantitative determination of biological activity is usually much less accurate and precise.

The choice of biological activity is in itself important. The data, normally for only one species, have to give a representation of the harm that a chemical will inflict on a whole ecosystem, as it is impossible to simulate even the simplest ecosystem for a toxicity test. There are many sources of biological data used in QSAR analyses, and these are extensively reviewed by Goldberg (1983). Devillers and Lipnick (1990) list 36 aquatic species that have been used in correlation analysis. These span several different trophic levels in the aquatic environment, from bacteria, blue-green algae, aquatic invertebrates, to a range of fish species. The biological endpoint most commonly utilised in environmental QSAR to produce quantitative (as opposed to qualitative) predictions is the concentration of a chemical that causes a specific biological effect. Especially popular and useful is the concentration that produce a 50% biological effect (although 0%, 10%, 90% and 100% have been used) e.g. LC_{50} is the concentration that kills 50% of the test species; EC_{50} is the concentration that produces a 50% reduction in a biological activity such as respiration in higher animals, or growth in lower

organisms. It may, however, be more appropriate for risk assessment in the environment, to be able to produce a value that gives an indication of the lowest concentration of a toxicant that will cause harm e.g. the minimum lethal dose, or the highest concentration that produces no effect - the no observed effect concentration.

Large data bases of toxicity values are available e.g. the Registry of Toxic Effects of Chemical Substances (RTECS), however, their use as a source of reliable data for QSAR purposes is notoriously difficult, as they tend only to collate the highest toxicity value, and results that may not be strictly comparable due to differences in methodology. All studies should be based on reliable data, with consistent methodology behind them. Such data have been provided by the U.S. E.P.A. for fathead minnow 96 hour LC_{50} and by Schultz and his co-workers for the 48 hour EC_{50} to Tetrahymena pyriformis growth. Both these can be considered reliable as they have been measured at the same respective laboratory. Other tests with high standardisation of procedure are the 48 hour LC_{50} of Daphnia magna and the 5 min EC_{50} to light production of Photobacterium phosphoreum (commercially known as the Microtox test).

1.3 Physico-Chemical Descriptors Utilised in QSAR Analysis

Since the modern re-birth of QSAR in the early 1960's, physico-chemical descriptors have been used to quantify three functions of a molecule:

i) its partitioning into a lipid biophase - hydrophobicity

ii) its size, shape, and/or symmetry - steric

iii) its reactivity and interactivity - electronic

The properties of a molecule which parameterise each of these attributes are described below.

1.3.1. Hydrophobicity

The hydrophobicity of a compound is a very important factor in determining its biological activity (Dearden 1985). The hydrophobicity ('water-hating') of a compound is the physical property of the molecule which governs its partitioning into a non-aqueous solvent. An increase in hydrophobicity thus generally results in an increase in transport of the xenobiotic into a lipid membrane, with a subsequent increase in biological activity (whether it be harmful, such as the inhibition of respiration or benefical such as an analgesic effect), this effect ceases however when chemicals are too lipophilic ('lipid-loving') to leave a lipid membrane and transport is reduced. Chemical properties commonly used as estimates of hydrophobicity include the partition coefficient, retention values from chromatography, and water solubility. Possibly the most important of these parameters is the partition coefficient, which may be simply defined as the equilibrium constant for the following process (Martin, 1978):

$$[\text{Drug}]_{\text{water}} = [\text{Drug}]_{\text{oil}} \qquad (1.1)$$

or

$$P = \frac{[\text{Drug}]_{\text{oil}}}{[\text{Drug}]_{\text{water}}} \qquad (1.2)$$

where P is the partition coefficient, the square brackets indicate concentration and the subscripts the phase

Early measurements of the partition coefficient used olive oil as the lipid phase (Meyer 1899; Overton 1899). This has now been superseded by the use largely of n-

octanol, although other solvents e.g. cyclohexane, diethyl ether and benzene are used. The measurement of partition coefficient has, however, been dogged by inaccuracies and variations in reported values resulting from variations in such factors as temperature, mutual phase saturation, pH, buffer type and concentration, and many more. These are fully discussed and evaluated by Dearden and Bresnen (1988), who recommend that partitioning should be carried out at constant temperature using either a stirred flask method technique or the filter probe.

Another commonly utilised estimate for hydrophobicity is by reversed phase, high performance liquid chromatography (RP-HPLC). Many studies studies have shown that log P is well correlated to capacity factors (k) measured on a reversed phase C18 column (Warne et al 1989).

Calculated estimates of log P are now commonplace in drug evaluation and QSAR studies. The first attempt in this direction was by Hansch et al (1962) who defined the hydrophobic substituent constant pi for any substituent X:

$$\underline{pi}_X = \log P_{YX} - \log P_Y \qquad (1.3)$$

where log P is the octanol/water partition coefficient
Y is any appropriate parent structure

pi values for a wide range of substituents have been tabulated by Hansch and Leo (1979). The use of pi is, of course, restricted to congeneric series, since it is a measure of substituent hydrophobicity.

A more general method to calculate log P was proposed by Nys and Rekker (1973). This was based on assigning 'fragmental constants' f to a variety of structural pieces, the calculated log P then being simply the sum of the f values appropriate to the molecule in question plus any interaction factors F that might be necessary to correct for intramolecular electronic, steric, or hydrogen-bonding interactions between fragments. By 1979, Rekker and de Kort had refined the fragment values by using a database of over 1000 log P measurements. They further proposed that all the interaction factors could be treated as multiples of 0.28 (the so-called 'magic number').

Hansch and Leo (1979) developed the fragmental scheme by carefully defining what constitues a fragment, and using a small basis set of carfully validated log P measurements, derived a value for each fragment. This model has been computerised as the ClogP program in the MEDCHEM software. It proceeds using a 'constructuralist approach', with hydrophobic constants being assigned to each fragment of a given molecule. The sum of these, together with correction factors, being used as an estimate of log P. Good correlations have been obtained for estimates of log P from MEDCHEM and measured values (Dearden, 1990), and the algorithm is now considered as the 'industry standard' and is widely used (Tute, 1990). (More information is given in section 2.2.1). In addition, the Rekker method of log P estimation has been computerised in the PrologP software available from Compudrug.

Other methods for calculating P have been proposed. Leahy (1986) parameterises log P in terms of a linear solvation energy relationship (LSER)

$$\log P = 5.14\ V/100 - 0.29\ u - 3.58\beta + 0.41 \qquad (1.4)$$

no statistics given

where V is the intrinsic molar volume
u is the dipole moment
β is the hydrogen bond acceptor basicity

Kamlet et al (1988) extended this approach by making it possible to estimate values for V, u, and β. This means log P is truely predictable by this method with high accuracy.

Using a similar approach Bodor et al (1989) have found a nonlinear regression model gives a good estimate of log P. The model is based on 13 molecular descriptors including measures of molecular surface, volume, weight, and MNDO calculated charge densities.

1.3.2.Descriptors of Electronic Effects

Electronic parameters are important descriptors of a chemical's reactivity and interactions with other molecules, which may control such phenomena as its binding to receptor sites, and/or its metabolism. The first major step in the development of electronic parameters took place in the 1930's, when Hammett (1935) proposed the Hammett sigma constant, σ_X, to assign numerical values for the electronic effect of substitution on an aromatic ring. With benzoic acid as the reference compound, this electronic parameter is defined by the equation:

$$\sigma_X = \log (K_X/K_H) \qquad (1.5)$$

where K_X and K_H are the ionisation constants for the x-substituted and unsubstituted benzoic acid, respectively.

Positive values of σ represent electron withdrawal by the substituent from the aromatic ring, and negative σ values indicate electron release to the ring. Taft (1956) extended Hammett's idea to aliphatic compounds. The effect of substitution on the reaction rates of aliphatic compounds is characterised by the parameter σ^*. This is defined as:

$$\sigma^* = \frac{1}{2.48} \left[\log \frac{k_X}{k_{H\,B}} - \log \frac{k_X}{k_{H\,A}}\right] \qquad (1.6)$$

where k_X is the rate constant for the hydrolysis of esters type $X\text{-}CH_2COOR$
k_H is the rate constant for the parent molecule
B refers to hydrolysis under basic conditions
A refers to hydrolysis under acid conditions

In the latter part of the 1950's interest began to develope in quantitatively separating the inductive (polar) part of the electronic effect of substituents from the resonance component. It had been presumed that it could be solved simply as:

$$\sigma = \sigma_I + \sigma_R \qquad (1.7)$$

where σ_I and σ_R represent the inductive and resonance components of the Hammett constant.

Although it was hoped that a single σ_R parameter would prove suitable for all correlation work, it became clear that 'through resonance' greatly complicates the picture. Four different types of σ_R can be considered (Dayal et al 1972;

Ehrenson et al 1973).

Swain and Lupton (1968) established another pair of substituent constants, F and R, for field-inductive and resonance effects separately. A more statistical approach was employed whereby they quantified the substituent effects in a bicyclooctane system as a measure of F, and it was assumed that R=0 for $p(CH_3)_3N^+$ on benzoic acid. F and R values were then calculated for 42 substituents and the importance of resonance (%R) in various σ constants was evaluated.

Hansch and Leo (1979) have tabulated σ, F, and R values for many substituents.

Measures of whole molecule reactivity have been utilsed in QSAR studies. Hermens et al (1985b) found that the reaction rate constants of reactive organic halides with 4-nitrobenzylpyridine (NBP) could be beneficial in describing their tendency to react with nucleophiles. Deneer et al (1987) suggested an alternative test with thiourea replacing NBP, for compounds exhibiting moderate or strong alkylating properties.

Electronic effects in a molecule are modelled by such properties as molar refractivity, pKa, dipole moment, NMR chemical shifts and a measure of hydrogen bonding (Dearden 1990). These descriptors are traditionally experimentally determined, and thus are of limited use in the modern QSAR environment. However, algorithms have been produced to calculate pKa (Hunter 1988; Gruber and Buss 1989), dipole moment from molecular orbital programs (see section 1.5.2), and molar refractivity using the MEDCHEM software (see section 2.2.1.5).

Accompanying the increase of the use of molecular modelling and computational chemistry software (see section 1.5) has been an increase in the number of electronic descriptors available for calculation. Molecular modelling is a very useful, if under used, tool in environmental science (Hauk and Schramm 1990). It is now possible to obtain reliable calculated estimates of whole molecule parameters such as HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) energies, and dipole moment. Also individual atomic descriptors can be obtained, such as charge, frontier electron density, and superdelocalisabilities. These are, of course, generally applicable only when

congeneric series of compounds are studied.

1.3.3. Steric Descriptors used in QSAR

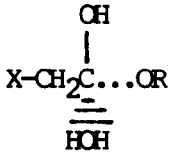
The size, shape, or bulk of the whole, or part of a molecule may be important in determining its activity if, for instance, it may not fit into a receptor site. The first generally successful numerical definition of steric effects in organic rections was that of Taft (1952), who defined the steric constant E_S as

$$E_S = \log \left(\frac{k_X}{k_h} \right)_A \qquad (1.8)$$

where k refers to the rate constant for the acid hydrolysis (denoted by A) of esters of type I:

$$X\text{-}CH_2COOR$$

The size of X will affect attainment of the transition state II:



by water in the case of acid hydrolysis or by some other ligand in the processes that are to be modelled by E_S.

Charton (1975) made a more direct approach to defining the steric hindrance of the substituent in the expression:

$$v_X = r_{vX} - r_{vH} \qquad (1.9)$$

where v_X is the Charton steric parameter for substituent X
r_{vX} is the minimum van der Waals radius for substituent X
r_{vH} is the van der Waals radius for hydrogen

Another set of steric substituent parameters has been developed by Verloop et al (1976) - the so-called Sterimol parameters. They undertook a multiparameter approach selecting 5 dimensions for each substituent and created a computer program using van der Waals radii, standard bond angles and lengths, and 'reasonable' conformations to define the space requirments of a molecule. The five dimensions define a box around the substituent and are labelled L, B_1, B_2, B_3, and B_4. The length parameter is defined as the length of the substituent

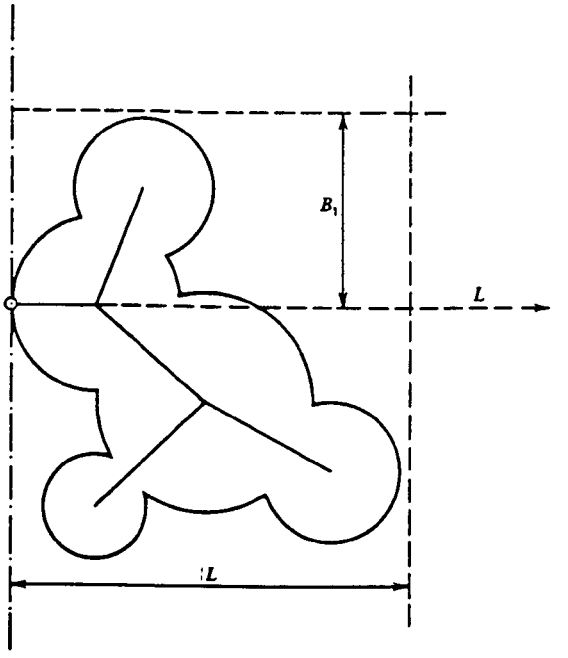
along the axis of the bond between the first atom of the substitution and the parent molecule. The four width parameters B_1-B_4 are determined by the distance at their maximum point perpendicular to this attachment bond axis and each other. B_1 is the smallest and B_4 is the largest width. Figure 1.1 shows a diagrammatic representation of these parameters. It must noted however, that the use of Sterimol parameters has almost exclusively been within pharmacuetical research. Steric descriptors which describe the size or bulk of the whole molecule can range from molecular weight and a simple count of carbon atoms to complex computer calculated estimates of the surface area of the molecule. Molecular weight and the number of carbon atoms have been used in QSAR studies; however, with the increasing use of computer technology to solve complex problems more sophisticated methods of calculating size are being produced. The algorithms written to calculate estimates of molecular size can be split into two main areas, those based on the topological content of a molecule, and those that use standard estimates of 'space' for each atom and attempt to represent the molecule by its atomic radii.

The most commonly utilised of all steric descriptors in environmental QSAR have been the topological indices. A topological index is a numerical descriptor of molecular size and is sensitive to features such as size, shape, symmetry and heterogeneity of atomic environments in the molecule. Two types of topological indices, connectivity and molecular complexity, are commonly used, although many topological indices have been proposed as chemical descriptors in QSAR. Balaban et al (1983) alone describes 27 such indices.

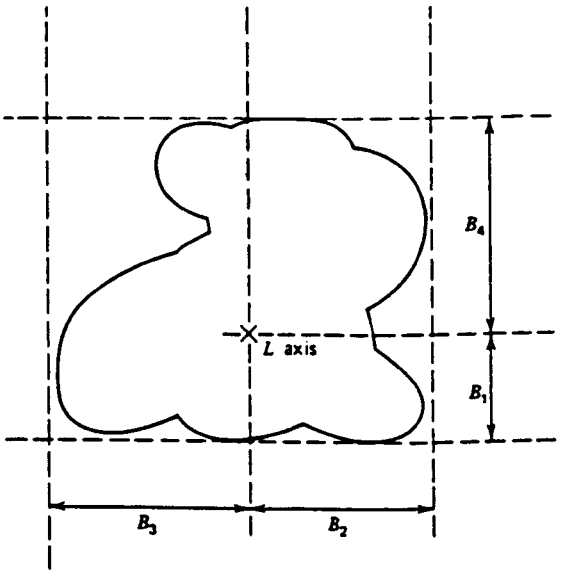
Of all the indices, molecular connectivities appear to be the most commonly incorporated into QSAR. (The derivation, calculation and use of molecular connectivities and Kappa indices are more fully described in section 2.2.2 and Appendix 2.) First devised by Randic (1975), they were greatly extended by Kier and Hall (1976). They encode the branching of a molecule and are calculated from a knowledge of atom connections within a molecule. Kier and Hall (1986)

Figure 1.1 Representation of Sterimol parameters as defined by Verloop et al (1976). (Figures taken from Hansch and Leo, 1979)

i) Perpendicular cross-section of the substituent along the L axis



ii) Vertical section viewed 'down' the L axis



extensively review their work and give numerous examples of their application in pharmaceutical and environmental QSAR, as well as their ability to predict chemical properties such as lipophilicity, molar refractivity, and Hammett constants. Because no exact physical nature has, or can be, attached to their value, use of molecular connectivities is still considered controversial and under evaluation. Dearden et al (1988) however do suggest that they primarily reflect the bulk properties of the molecule. Kier (1985, 1987) has also developed the Kappa indices, from the number of atomic fragments in a molecule, which are thought to quantify an element of molecular shape or symmetry. Less commonly used are the molecular complexity indices, with Information Content (Basak and Magnuson 1983), Structural Information Content, and Complementary Information Content (Basak et al 1980) proving most popular and reliable as QSAR parameters. The estimates of surface area and molecular volume are calculated from an energy minimised structure of the molecule in question. The 'accessible' surface is calculated from the van der Waals radius of each atom, and obtained by simulating the rolling of a probe (of defined radius) over it. Thus to obtain a value for the total surface area a probe radius 0.0 is used, and to measure how much of the molecule would be open to a substrate of a given diameter (e.g. a water molecule) different probe radii can be used. Measures of accessible surface area are thought to give a better representation of molecular bulk than simple surface area calculations as they are more likely to take account of atomic overlap and conformation (Dearden, 1990). It is also thought that the accessible surface area of a molecule is important in determing it's hydrophobicity, e.g. de Bruijn and Hermens (1990) found a highly significant relationship between the log P for polychlorinated biphenyls and the solvent accessible surface area (SASA):

$$\log P = 0.027\ SASA - 7.12 \qquad (1.10)$$

$n=18 \qquad s=0.12 \qquad r^2=0.986 \qquad F=1116$

1.4. Statistical Techniques used in QSAR Analysis

It is the statistical technique that 'ties' together the physico-chemical and biological data in a QSAR analysis. Careful application of statistical methodology will reveal much more information from the data than is initially apparent. A wide variety of techniques has been used to develope QSAR. The most commonly used methods are described below.

1.4.1. Regression Analysis

The most popular technique in QSAR has been regression analysis. It has a number of distinct advantages, being predictive and easy to understand, as well as being able to highlight outliers from a particular relationship. Disadvantages include its need of quantitative biological data, the problem of chance correlations and collinearity between the independent variables, and its reliance on continuous parameters. A regression analysis depicting the linear relationship between X (the dependent variable) and Y (the independent variable) has the general form:

$$X = aY + b \qquad (1.11)$$

where a is the slope of the regression line
b is the intercept

Historically it was Hansch and Fujita (1964) and Free and Wilson (1964) who developed two different techniques for deriving QSARs, each essentially based on regression analysis. The Hansch method has been particularly widely used.
In the early sixties Hansch et al (1963) postulated that biological response is a linear function of one or more of three related main properties, hydrophobic, electronic and steric. The influence of hydrophobicity is related to the probability of a drug or toxicant reaching the site of action. The influence of electronic and steric factors, as well as hydrophobicity, can be involved in the xenobiotic reacting or interacting with the receptor. Electronic factors also can influence the degree of ionisation of chemicals, and hence, the rate of uptake, since generally only undissociated molecules can penetrate lipid membranes. Usually the Hansch approach is applied to congeneric series of compounds. The hydrophobic effect is characterised by the n-octanol/ water partition coefficient

(expressed as log P), and electronic and steric effects can be modelled by, for example, the Hammett σ constant and the Taft steric constant respectively. Hansch et al (1963) derived the following equation to describe the relation between biological activity and physicochemical properties for a set of compounds

$$\log 1/C = a (\log P)^2 + b (\log P) + c\sigma + dE_s + e \qquad (1.12)$$

where C represents the molar concentration that elicits a constant biological response from a given organism (e.g. ED50)
σ is an electronic term
E_s is a steric term

The constants can be derived by multiple linear regression.

The goodness of fit is expressed by statistical criteria such as the correlation coefficient (r) and the standard error (s). Recently, however, it has become accepted that the coefficient of determination adjusted for the number of degrees of freedom of the equation (r^2(adj)) should be used (Moulton 1988). This indicates how much of the variance of the dependent variable is explained by the independent variables. Also required is the F value, called the Fisher statistic, which allows estimation of whether the obtained relationship is statistically significant. In addition, in this study, all calculated regression equations are given with the standard error of the coefficients of each variable shown in parentheses.

The r^2, s and F values are used as initial tests (before validation) to decide whether the relationship is strong or weak. A strong relationship would be expected to have a high r^2 (preferably over 0.9), low s, and a highly significant F value. It can then be concluded that such a QSAR has modelled the relationship well, and with careful use it may provide useful predictions and information. These suppositions must, of course, be supported by thorough validation of the model. Weaker QSARs i.e. those with a low r^2, high s, and a less significant F value have consequently been less successful in modelling the relationship and their interpretion and use should be treated with more caution.

Present day Hansch-type QSAR employs stepwise and best-subsets regression analysis to find the best equations to predict biological activity if many

physico-chemical data are used.

1.4.2. Free-Wilson Technique

This method (Free and Wilson 1964) is based on the assumption that in a series of related chemicals each particular substituent adds a constant contribution to the biological activity (BA) of the molecule. Thus

$$f(BA) = Sa + Sb \ldots + u \qquad (1.13)$$

where u is the contribution of a hypothetical parent compound to the biological activity
Sa and Sb are the contributions added by groups A, B etc.

Each compound yields an equation of the type shown above and the group contributions are found by solving a set of multiple simultaneous equations (Purcell et al 1973). The 'goodness of fit' or 'correlation' can be determined by regression analysis as in other QSAR methods.

An advantage of the Free-Wilson approach is that physico-chemical or other properties need not be determined; they are contained within the additive group contributions. Predictions are limited to compounds comprising the parent molecule and the substituents in the training set. The numerical values of the group contribution will depend upon the method of expressing the biological activity. Because specific physico-chemical properties are not ascribed to the substituent groups, information on the mechanism of biological action is not usually obtained by this approach.

1.4.3. Multivariate Data Space and Statistical Analysis

The absorption of disciplines such as computer chemistry into established QSAR practice has led to a common scenario whereby there are many more physico-chemcial (independent) data than biological (dependent) data. Such a data matrix has been termed 'over-square' (Hyde and Livingstone 1988), Fig 1.2 shows a diagrammatic representation of a matrix. Several problems arise from the statistical manipulation of over-square matrices. There is a considerable redundancy of information; i.e. many descriptors will represent the same molecular feature, leading to collinearity within the data. A collinear data set may lead to chance correlations, giving spurious results if used in techniques

Fig. 1.2 Diagrammatic Representation of a Typical Multivariate QSAR Data Matrix (Adapted from Dunn 1988).

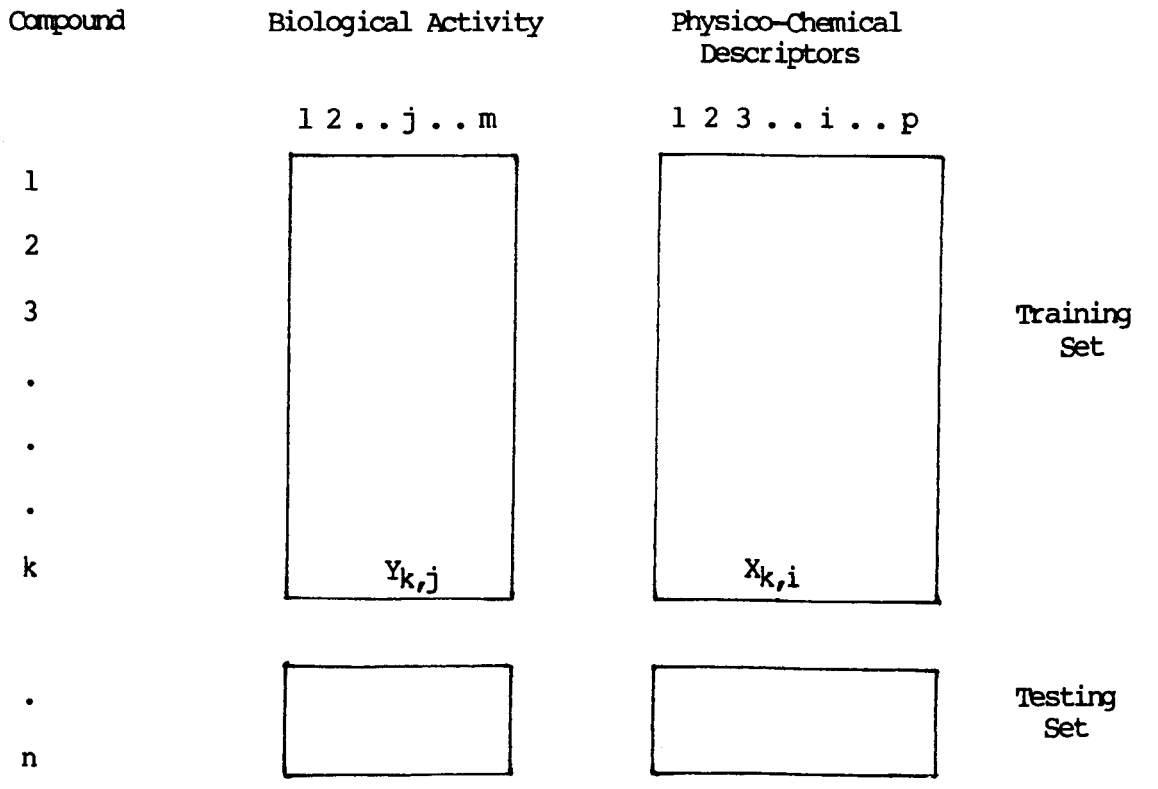


Table 1.1 Principal Multivariate Statistical Methods used in QSAR Analysis

(adapted from Hyde and Livingstone 1988)

Supervised Learning

1. Multiple regression
2. Discriminant analysis
3. Linear learning e.g. PLS, SIMCA
4. Canonical correlation
5. Adaptive least squares

Unsupervised Learning

6. Non-linear mapping
7. Principal components
8. Factor analysis
9. Cluster analysis
   a) on variables
   b) on cases
   c) k-nearest neighbour
10. Correspondence analysis

such as stepwise regression analysis (Kikushi 1987). In an attempt to rectify this problem there has been an increase in the use of multivariate statistical methods (Livingstone 1989). The main techniques are summarised in Table 1.1. Methods can be split into two areas, supervised and unsupervised learning. The supervised learning methods rely on biological activity to be present so that the model may be formed. Unsupervised methods, however, do not need any form of class membership to be assigned for a model to be created.

To appreciate the action of multivariate statistical methods, the data matrix should be considered as a multi-dimensional entry into space. For n parameters there are n dimensions. The data for the compounds are projected into the hyperspace so that each one is represented by a discrete point. Some multivariate techniques, broadly referred to as pattern recognition, e.g. cluster analysis, k-nearest neighbour, discriminant analysis, and SIMCA utilise these properties in the hyperspace to enable a qualitative distinction to be made amongst the compounds (Wold and Dunn 1983). This is especially beneficial when activity is classified categorically, typical examples being active and non-active drugs, carcinogenic and non-carcinogenic compounds etc. Thus for a compound with unknown activity, its position in the hyperspace relative to compounds with known activity, may mean that an estimation of its activity can be obtained.

Other multivariate techniques attempt to reduce collinearity in tha data without significant loss in its information content. Cluster analysis on variables will for instance place the physico-chemical data themselves in an n-dimensional framework and form clusters of the data points, according to which is closest. Thus an immediate reduction in data is achieved by simply discounting all but one, or a few members of each cluster. (The full significance of this method is discussed in section 4.2.1.)

Methods such as principal component analysis (PCA) and factor analysis (FA) calculate new orthogonal variables from the data, each accounting for a particular feature of those data. Niemi (1990) lists three objectives for using

these methods:

i) to explore and detect patterns among a set of variables, in association with data reduction;

ii) confirmatory testing of the underlying 'structure' among a set of variables and the relative factor loadings for the variables;

iii) to develop new variables to serve as new, simpler parameters in subsequent analysis.

This is achieved, after the preparation of the data matrix, by the isolation and extraction of the initial principal components (which includes the data reduction phase), and in factor analysis additional calculations, such as rotation of the factors in the n dimensional space, are undertaken to aid the interpretation of simpler factors.

Canonical correlation analysis (CCA) seeks to form a relationship between two data sets for which more than one variable exists in both sets. It can be seen as an extension of multiple regression analysis (where one of the two data sets contains only one variable), although some authors (e.g. Lindeman et al, 1980) have observed that it might be more appropriate to view multiple regression as a special case of canonical correlation. It proceeds by the formation of new orthogonal variables (known as canonical variables) for both data sets, and a maximisation of the correlation between the new variables. The interset association between the canonical variables is termed the canonical correlation coefficient. This technique can be used in QSAR analysis when both the biological and descriptor data sets are multivariate, to elucidate more about the potency, or mode of action of a chemical. For instance, Szydlo et al (1984) have assessed the relationship between the knockdown activity and toxicity of a series of pyrethroid insecticides.

1.4.4. Statistical Validation of Models

It is essential to assess the validity of a statistical model so that its predictive strength and robustness can be realised. The r^2(adj), s, etc statistics show how well the model fits the data, yet tell us little about how

accuarate it is in predicting an unknown biological activity. In early environmental QSAR there was little evidence of statistical validation being carried out. However two separate methods, both of merit, are increasingly being applied.

The first, and possibly more pertinent, method of validation is that of creating two data sets within the data. The first is the 'training set' on which all the quantitative modelling is based. The second is a 'test set' of data not included in the model. The precision of the model can then be tested by calculating the biological activity of the test set and comparison with the measured activity. Tosato et al (1990) review the procedure for the creation of both data sets.

The second method of validation is referred to variously as cross-validation, jackkinfe, bootstrapping, or simply leave-one-out. For a data set the model is recalculated after one (or a group of) compounds has been left out, and is this reapplied to that compound (or group of compounds) to estimate the biological activity. This is repeated until each one (or group) has been left out in turn. The deviation from the expected results can then be evaluated (Gray and Schucany 1972).

1.5. Computer Chemistry and Molecular Graphics

In the 1980's a vast increase in the availability of computing power has led to the expansion in the use of computer chemistry in the area of drug design. Computer chemistry is the quantitative modelling of chemical behaviour on a computer by the formalisms of theoretical chemistry. Allied with the growth in molecular graphics it proves a valuble tool in the estimation of a molecule's three dimensional properties (Hopfinger, 1985). It is also a remarkable asset for the scientist as it allows calculations to be performed on compounds regardless of whether or not they have been synthesised. As well, information is provided in a fraction of the time, and cost, of that determined experimentally. The combined techniques of computer chemistry, molecular graphics, together with the growth of computer technology has created a new area of investigation - loosly termed Computer Aided Drug Design (CADD) or Computer Aided Molecular Design (CAMD). This has yielded many publications and more recently its own journal - Journal of Computer Aided Molecular Design.

The assessment of molecular structure through molecular design begins by entry of the structure into a program either as a set of cartesian (XYZ) coordinates, manual graphical entry of individual atoms, substructures, or fragments using a 'mouse' or bit-pad in an appropriate entry environment, or by a code e.g. SMILES. Calculations are then performed on the chemical structure in order to obtain its theoretically most stable conformation i.e. that with the lowest electronic energy. Methods of achieving this, the 'minimisation' of the molecule, are broadly split into two areas - the molecular mechanics and molceular orbital approaches.

1.5.1.1. Molecular Mechanics

Molecular mechanics, or force field calculations are based on a simple classical-mechanical model of molecular structure. The molecule is considered as a set of balls (atoms) held together by springs (the potential functions of the bond lengths and angles) (White, 1977). The first step in the molecular mechanics calculation is the determination of interatomic distances, bond angles, and

torsional angles in the starting geometry. The values obatined are then used in one of the many different potential function expressions to calculate an initial steric energy (E). Steric energy is specific to its force field, and as such has no physical meaning. It is approximated as the sum of the energy contributions

$$E = E_s + E_b + E_w + E_{nb} + \ldots.$$

where E_s is the energy of a bond being stretched or compressed
E_b is the energy of bending bonds
E_w is the torsional energy due to twisting about bonds
E_{nb} is the energy of nonbonded interactions

Other intramolecular mechanisms affecting the energy, such as electrostatic (coulombic) repulsions or hydrogen bonding may be added to the force field (Boyd and Kipkowitz, 1982).

1.5.1.2. Example of a Potential Energy Function - Bond Length

The principle of the potential energy function is best considered using the bond-stretching term as an example. The potential at any interatomic distance r is described by the Morse curve, with a maximum at r_o (see Fig. 1.3). The expression for a Morse curve is, however, complicated and requires too much computer time to solve. However, since the vast majority of molecules have bond lengths within a limited area, symbolised by the shaded portion in Fig. 1.4, Hooke's Law gives a good fit to the energy profile, expressed by

$$V = k(r-r_o)^2/2$$

where V is the potential energy
k is a constant

1.5.1.3. Geometry Optimisation

Once an initial energy has been obtained the process is continued with the aim of revealing the minimum energy geometry. This is performed by optimisation techniques, such as the Newton-Raphson method, which use analytically evaluated derivatives of molecular energy. Such methods work by the 'finite difference' method. Considering bond length as an example, its energy potential curve is displayed in Fig. 1.4. The energy of the initial geometry is calculated, and then recalculated after a slight change in the geometry. The energy difference δE is

Figure 1.3 The potential energy curve for stretching a chemical bond. The dashed curve represents a simple Hooke's law potential-function.

(Taken from Clarke, 1985)

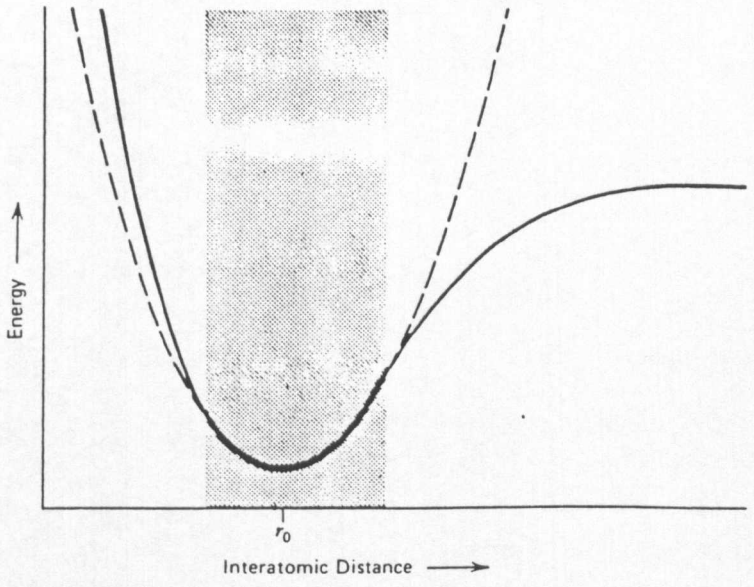
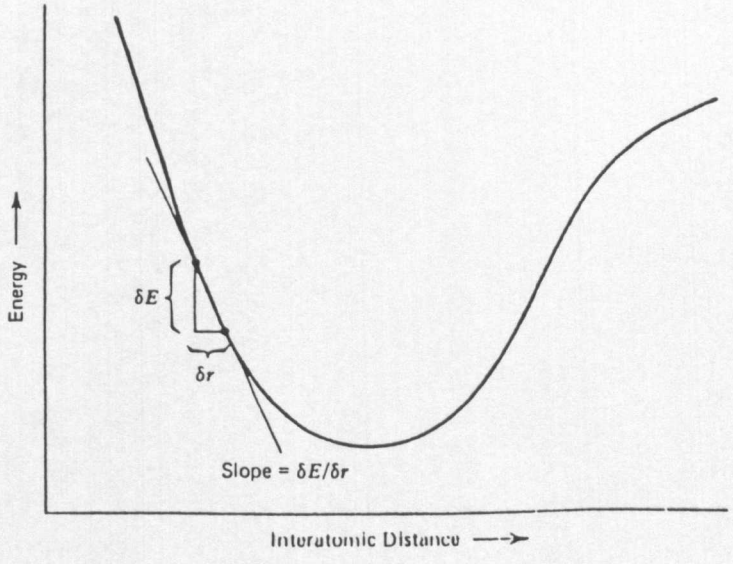


Figure 1.4 Determination of atomic forces by the finite difference method.

(Taken from Clarke, 1985)



used to determine the gradient $\delta E/\delta r$. The distance from the minimum should be proportional to the gradient, and the geometry is altered to obtain the next structure which should be closer to the minimum.

1.5.2. Molecular Orbital (MO) Theory

Molecular Orbital (or quantum mechanics) theory represents the molecule as a set of molecular orbitals to be occupied by the electrons assigned. The orbitals with the lowest energy are then sought. The commonly used semiempirical theory utilises the LCAO-SCF procedure. The LCAO (linear combination of the atomic orbitals) formalism describes every molecular orbital as a linear combination of the atomic orbitals using the following equation:

$$MO = X_1C_1 + X_2C_2 + X_3C_3 \ldots + X_nC_n$$

where MO is the molecular orbital
X_n is the n^{th} atomic orbital
C_n is the n^{th} coefficient of atomic orbitals' eigenvector

The SCF, the self consistent field operation, allows for the interactions between different electrons and orbitals. The LCAO equation produces the optimised bond lengths, bond angles, dihedral angles, free valences, the electron populations, atomic charges and all the other data of the electronic structure of a molecule (Clarke, 1985).

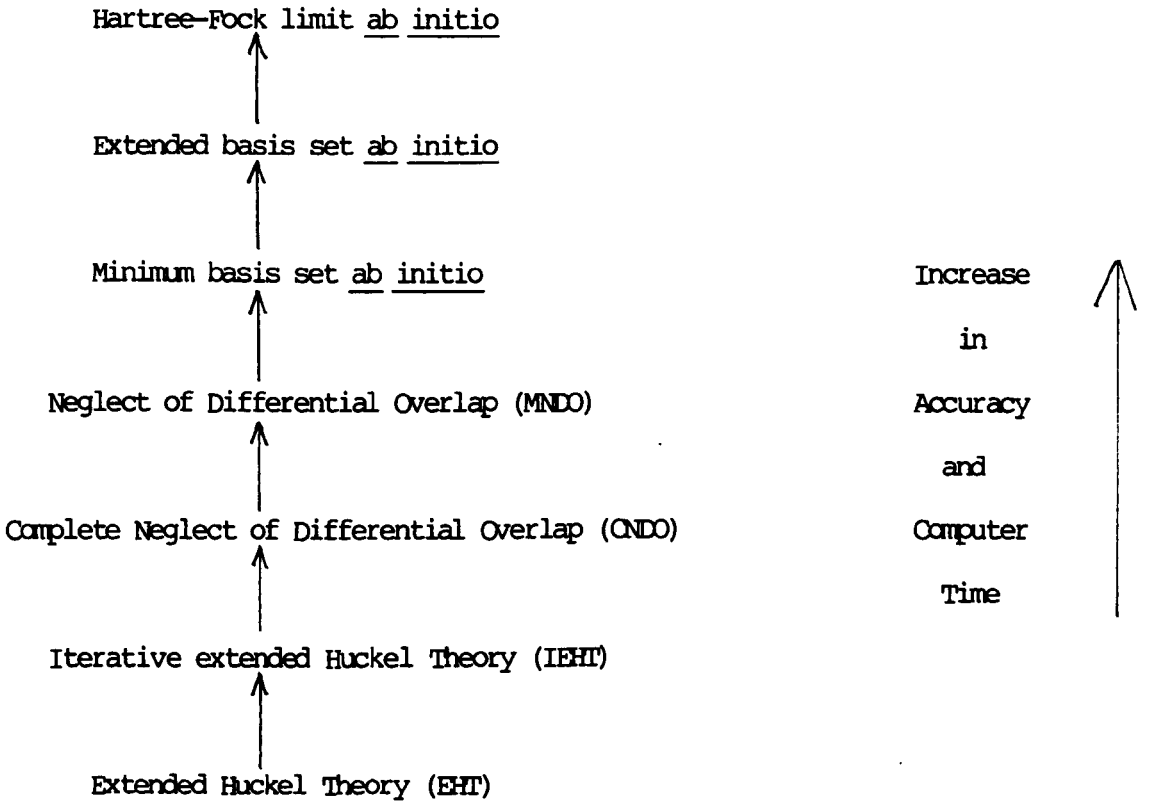
There are many types of MO program (see Fig 1.5). The ab initio methods are considered to be most accurate; however, due to their complicated nature and the fact that they require a lot of CPU time, they are not normally performed. Popular, however, are the neglect-of-overlap methods (CNDO and MNDO).

1.5.3. Computer Chemistry in QSAR Analysis

The ability to calculate the active conformation of a molecule is a most valuable tool for qualitative analysis e.g the elucidation of receptor shape. For quantitative analysis of molecules the calculation of wave functions from MO theory allows derivation of much information about the molecule. (An analysis is commonly performed by using molecular mechanics to optimise geometry, and MO theory to calculate charge distributions.) For instance, charge distributions can be derived from the square of the wave function at a point in space and

Fig 1.5. Summary of the Various Approximate Molecular Orbital Methods

(adapted from Richards 1983)



integrated over a defined volume. From this basis values can be obtained for the eigenvalue of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO).

Other atomic descriptoprs of possible use in QSAR studies have been devised. The concept of superdelocalisability was introduced by Fukui et al (1954). Denoting the occupied molecular orbitals by 1, 2, ... m, and the unoccupied levels by m+1, m+2, ..., the superdelocalisability, S_r is given for the three types of reaction by:

i) for an electrophilic reaction

$$S_r^{(E)} = 2 \sum_{j=1}^{m} \frac{C^2_{rj}}{\lambda_j}$$

ii) for a nucleophilic reaction

$$S_r^{(N)} = 2 \sum_{j=m+1}^{N} \frac{C^2_{rj}}{\lambda_j}$$

iii) for a radical reaction

$$S_r^{(R)} = \sum_{j=1}^{m} \frac{C^2_{rj}}{\lambda_j} + \sum_{j=m+1}^{N} \frac{C^2_{rj}}{-\lambda_j}$$

where c_{rj} is the coefficient of the r^{th} atomic orbital in the j^{th} molecular orbital

λ_i is the coefficient in the orbital energy, given by

$$E_j = \alpha + \lambda_j \beta$$

where α is the ionisation potential

β is an empirical energy parameter

Other features that can be calculated with a knowledge of the frontier electron densities include dipole moments, moments of inertia, and the principal ellipsoid axes of the molecule. In addition, estimates of the surface area can made from the 'optimum geometry'.

Parameters calculated from computer chemistry have already been shown to be valuable in environmental QSAR (Dearden and Nicholson, 1987; Purdy, 1988). They are divided into two types, descriptors of the whole molecule, and those of individual atoms in a molecule. Atomic descriptors, e.g. charges, superdelocalisabilities, are useful only when studying the effect of substitution on a parent compound.

Complete molecular modelling packages are now available which will find minimum energy structures and automatically calculate physicochemical descriptors from a simple input. Commercially available software such as Chem-X from Chemical Design Ltd, SYBIL from Tripos, and CHARMm from Polygen are good examples. Furthermore, the pharmacuetical industry has developed its own software, tailoring it to their needs, e.g. COSMIC from Smith-Kline and Beecham (Vinter, 1987), and Profiles from the Wellcome Corporation (Glen and Rose, 1987).

1.6.1. Historical QSAR

The biological activity of a chemical relies ultimately on its chemical structure. The use of Quantitative Structure-Activity Relationships (QSAR) in environmental toxicology is an attempt statistically to justify this fundamental principle (Turner et al, 1987). The basis of all QSAR can be traced back over 125 years to studies conducted on the effects of the so-called 'narcotic' chemicals. Cros (1863) determined the toxicity of methyl and amyl alcohols and found that they exhibited similar effects of depression followed by death at higher doses, regardless of the test species, or the route of administration. He also noted a decrease in water solubility with the addition of methyl groups to a molecule. It was, however, Crum-Brown and Fraser (1868) who first proposed that

> 'there can be no reasonable doubt but that a relation exists between the physiologic action of a substance and its chemical composition and constitution'

They also put forward a mathematical concept, thus

$BA = f(c)$

where BA is the biological activity
c is the 'constitution' of the chemical
f is a constant

The most famous, and possibly most valuable, early breakthroughs, however, were those of Overton (1899) and Meyer (1899). They independently proposed that the toxicity of simple non-electrolyte organic compounds depends on their ability to partition from water to a lipoid biophase site of action. They also suggested the use of the olive oil/ water partition coefficient as a model parameter. Hansch and Dunn (1972) used the data of some the compounds tested by Overton to derive a QSAR:

$$\log (1/C) = 0.901 \log P + 0.909 \qquad (1.14)$$

n=57 s=0.312 r=0.962 F not given

where C is the lowest concentration in moles/l found to produce narcosis in tadpoles

As already mentioned in Section 1.3.2. it was the work of Hammett in the 1930s

and Taft in the 1950s who progressed the art of SAR. Undoubtedly, however, the modern science of QSAR was founded by Corwin Hansch and his co-workers. Hansch et al (1962) quantitatively related the herbicidal activity of phenoxyacetic acids to Hammett substituent constants (σ) and the octanol-water partition coefficient (pi):

$$\log 1/C = 4.08\ pi - 2.14\ pi^2 + 2.78\ \sigma + 3.36 \qquad (1.15)$$

$n=20$

where C is the concentration inducing a 10% growth in Avena coleoptiles in 24 hours

These ideas have been constantly reviewed and developed, finding numerous applications in the design of new pesticides and drugs. Since QSAR is used to predict the beneficial effects of chemicals, it seemed logical that it can be used to predict the harmful effects of chemicals. Hansch and Dunn (1972) reported linear QSAR equations, of good predictive power, containing solely the log P parameter, for the 24 hour minimum lethal concentration of organic compounds to carp, goldfish, goby, roach, and tench fish, based on the 1901 data of Cololian. In 1974 Kopperman et al related the toxicity of a series of phenols to Daphnia magna using free energy terms. In addition, Ljublina and Filov (1975) describe early attempts to quantify toxicity, that were made in the USSR. These were probably the first publications in what is now being termed environmental QSAR; most of the work, however, has taken place in the 1980s. Research has since been concentrated in three main areas, the prediction of acute toxicity to aquatic species, the bioconcentration of chemicals in the environment, and to a lesser extent, the biodegradation of chemicals in the environment.

1.6.2. Modern QSAR - The Importance of Hydrophobicity

The historical importance of hydrophobicity in QSAR has already been emphasised by the work of Cros (1863), Meyer (1899), Overton (1899), and Hansch and Dunn (1972). In 1981 Konemann continuing this avenue of thought, found that the logarithm of the partition coefficient was an extremely good model of the acute toxicity to the guppy of 50 non-reactive, non-ionised organic chemicals, acting

solely by a narcosis mechanism: (see Fig. 1.6)

$$\log 1/LC_{50} = 0.871 \log P - 4.87 \qquad (1.16)$$

n=50 s=0.237 r=0.988 F not given

where LC_{50} is the molar concentration that causes 50% fish mortality in the guppy
log P is the calculated (according to the Rekker (1977) method) logarithm of the octanol/ water partition coefficient

Although outliers were found to this equation these were attributed to 'excess toxicity' i.e. the chemicals were acting by a more specific mechanism. Konemann concluded that the lethal effect was probably caused by membrane perturbation and seemed to be a kind of minimum effect i.e. a hydrophobic substance was at least as toxic as calculated from the QSAR unless it was strongly metabolised. This relationship ended however when chemicals had a log P greater than 6, due to insufficient water solubility.

This relationship between hydrophobicity and the acute toxicity of non-reactive chemicals has been seen many times elsewhere. For example, Veith et al (1983) found an extremely good relationship with the 96 hour LC_{50} of 60 common narcotic chemicals to the fathead minnow Pimephales promelas:

$$\log LC_{50} = - 0.94 \log P + 0.94 \log (0.000068P + 1) - 1.25 \qquad (1.17)$$

n=60 s, r, and F not given in text

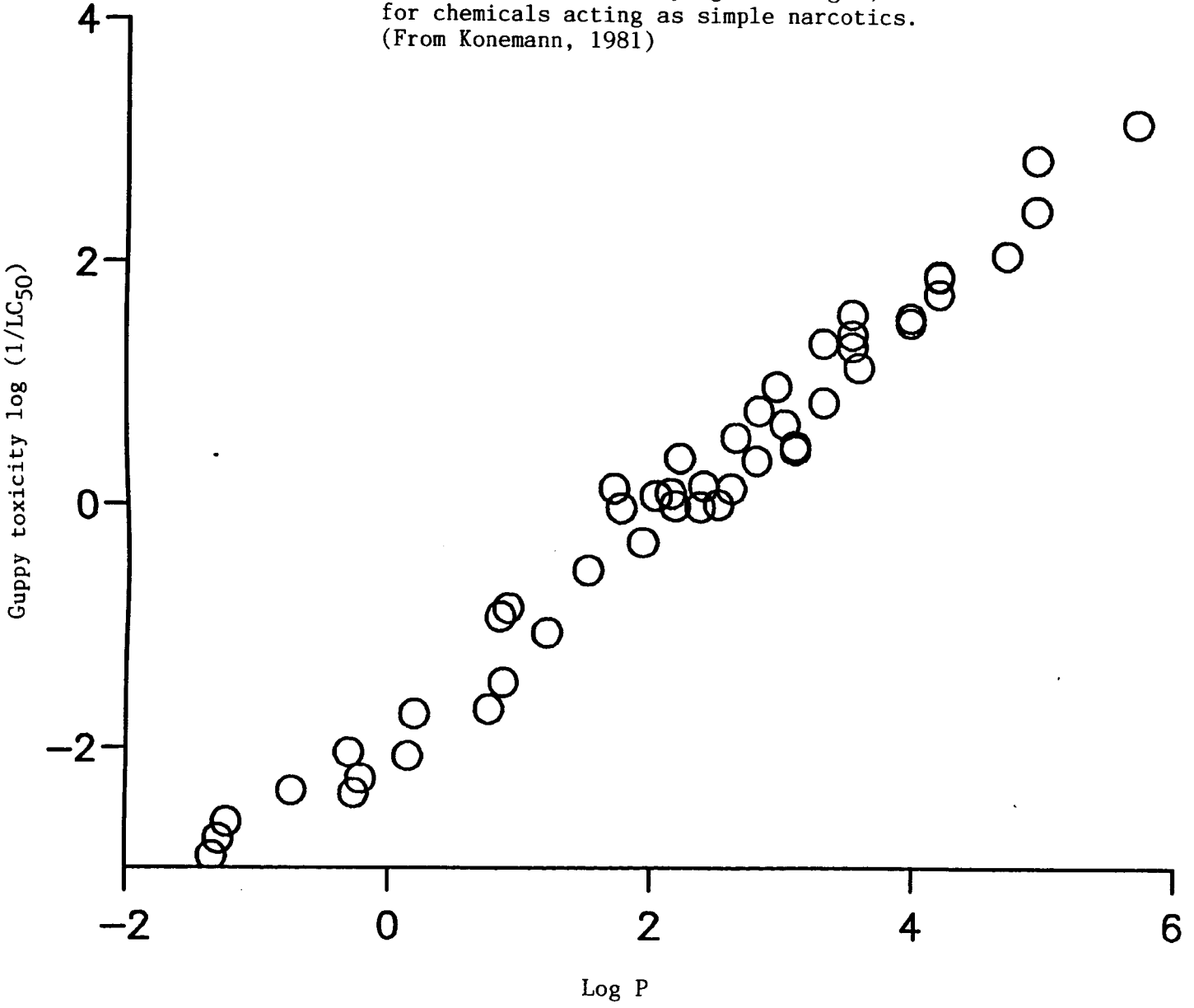
The toxicity of such unspecific toxicants has been described by other features of a chemical's partitioning, such as the capacity factor from high performance liquid chromatography (HPLC). Warne et al (1989) reports that capacity factors obtained on C-18 stationary phases, are the best descriptors of the toxicity of diverse organic compounds to a mixed marine bacterial culture, when compared with six other stationary phases:

$$\log EC_{50} = - 1.25 \log k(C18) - 0.006 BP + 3.491 \qquad (1.18)$$

n=17 s not given r^2=0.987 F not given

where $\log EC_{50}$ is the molar concentration causing a 50% decrease in growth after 16 hours in a culture of mixed marine bacteria
log k(C18) is the logarithm of the capacity factor of the C18 stationary phase for each chemical
BP is the boiling point

Figure 1.6 Graph of guppy toxicity against log P, for chemicals acting as simple narcotics. (From Konemann, 1981)



Protic and Sabljic (1989) discovered a strong relationship between the toxicity of eight classes of narcotic chemical to the fathead minnow and their zero order valence molecular connectivity indices (PV0):

$$\log 1/LC_{50} = 3.43\ PV0 - 0.68 \qquad (1.19)$$

n=113 s=0.541 r^2(adj)=0.84 F=568

where LC_{50} is the logarithm of the 96 hour molar LC_{50} to the fathead minnow Table 1.2 shows many more examples of partition coefficient having been used to model nonspecific narcosis.

This phenomenon of the Log P relationship has been termed "baseline toxicity" by Lipnick and Dunn (1983). They considered that narcosis or physical toxicity resulted from a simple, reversible, physicochemical process in which the biological response is solely a function of the molar cellular concentration of the toxicant.

1.6.3. Modelling Toxicity Mechanisms other than Narcosis

Many classes of chemical do not act solely as narcotic agents, i.e. they have a much more specific toxicity mechanism. This may arise from metabolism to a more toxic product, or because they may react irreversibly with an enzyme. The net result is that the chemical has a toxicity value greater than that predicted by a baseline narcosis hydrophobicity model. Some mechanisms of toxic action are described below.

1.6.3.1. Electrophile Toxicity Mechanism

Organic nonelectrolytes capable of reacting with sulphhydryl groups and other nucleophilic moieties present in target biological macromolecules may be classified in general as acting by an electrophile molecular mechanism (Lipnick and Dunn, 1983). Electrophile toxicants such as epoxides, allyl and benzyl chlorides, 2,4-dinitrofluorobenzene and chlorotriazine can undergo nucleophilic substitution reaction. Others, including alpha- and beta- unsaturated aldehydes, ketones, esters, sulphones, nitriles, and amides, can serve as Michael-type acceptors (Lipnick, 1989).

Hermens et al (1985b) attempted to model the 14 day LC_{50} for the guppy of 15

Table 1.2 QSAR Studies with Relatively Unreactive, Non-Ionised Organic Chemicals

(adapted from Hermens, 1986)

Effect and Species	Types of Compounds Tested *	Number of Compounds	References
LC50 to golden orfe	Not Exactly Given	89	Lipnick and Dunn 1983
LC50 to 14 aquatic species	2, 6, 9, 10	5	Sloof et al 1983
LC50 to zebra fish and rainbow trout	1	5	Calamari et al 1983
LC50 to guppy	1, 5, 6	17	Koch 1982
LC50 to bleak and harpacticoid copepod	9	14	Bengtsson et al 1984
Sublethal toxicity to guppy	9	9	Kier and Hall 1982
Sublethal toxicity to fathead minnow	1, 3, 5, 10, 11	10	Call et al 1985
Immobilisation of Daphnia magna	1, 2, 4, 7, 8	33	Bobra et al 1983
Immobilisation of Daphnia magna	1	5	Bobra et al 1985
Immobilisation of Daphnia magna	1-3, 5, 6, 9-11	19	Hermens et al 1984
Sublethal toxicity to Daphnia magna	1	6	Calamari et al 1983
Sublethal toxicity to Daphnia magna	1, 3	5	Hermens et al 1984
Sublethal toxicity to Daphnia magna	1, 3, 5, 9, 10	10	Hermens et al 1985
Toxicity to algae	1	6	Calamari et al 1983
Bioluminescence of bacteria	1, 2	11	Ribo and Kaiser 1983
Bioluminescence of bacteria	1-3, 5, 6, 9-11	22	Hermens et al 1985
Acute toxicity to ciliate protozoa	2, 4, 7, 8,	17	Rogerson et al 1983

* Types of chemicals: 1.Chlorobenzenes 2.Alkylbenzenes 3.Chlorinated alkylbenzenes 4.Alkanes 5.Chloroalkanes 6.Chloroalkenes 7.Biphenyls 8.Polycyclic aromatic hydrocarbons 9.Alcohols 10.Ketones 11.Ethers

reactive organic halides (acting as electrophiles). The toxicity was only poorly modelled by log P, with many toxicities considerably greater than predicted by baseline narcosis: (See Fig 1.7)

$$\log 1/LC_{50} = 0.474 \log P - 1.98 \qquad (1.20)$$

n=15 s=1.11 r=0.41 F not given

However the QSAR was greatly improved by the addition of a new parameter (K), the reaction rate constant of the chemicals with 4-(4-nitrobenzyl)pyridine (NBP). This is well recognised as a parameter for chemical reactivity in nucleophillic substitution reactions (Panthananickal et al, 1978):

$$\log 1/LC_{50} = 0.224 \log P - 1.32 \log(2484 + K^{-1}) + 4.05 \qquad (1.21)$$

n=15 s=0.39 r=0.956 F not given

Deneer (1988), one of Hermens's coworkers, has found many more such relationships, and has proposed (Deneer et al, 1987) the use of thiourea reaction rates as an alternative to NBP for moderate and strong alkylating agents. Lipnick et al (1987) has termed toxicity greater than that predicted from baseline narcosis theory as 'excess toxicity' (T_e). This may be defined as the ratio of predicted to observed toxicity. i.e.

$$T_e = \frac{LC_{50}(pred)}{LC_{50}(obs)} \qquad (1.22)$$

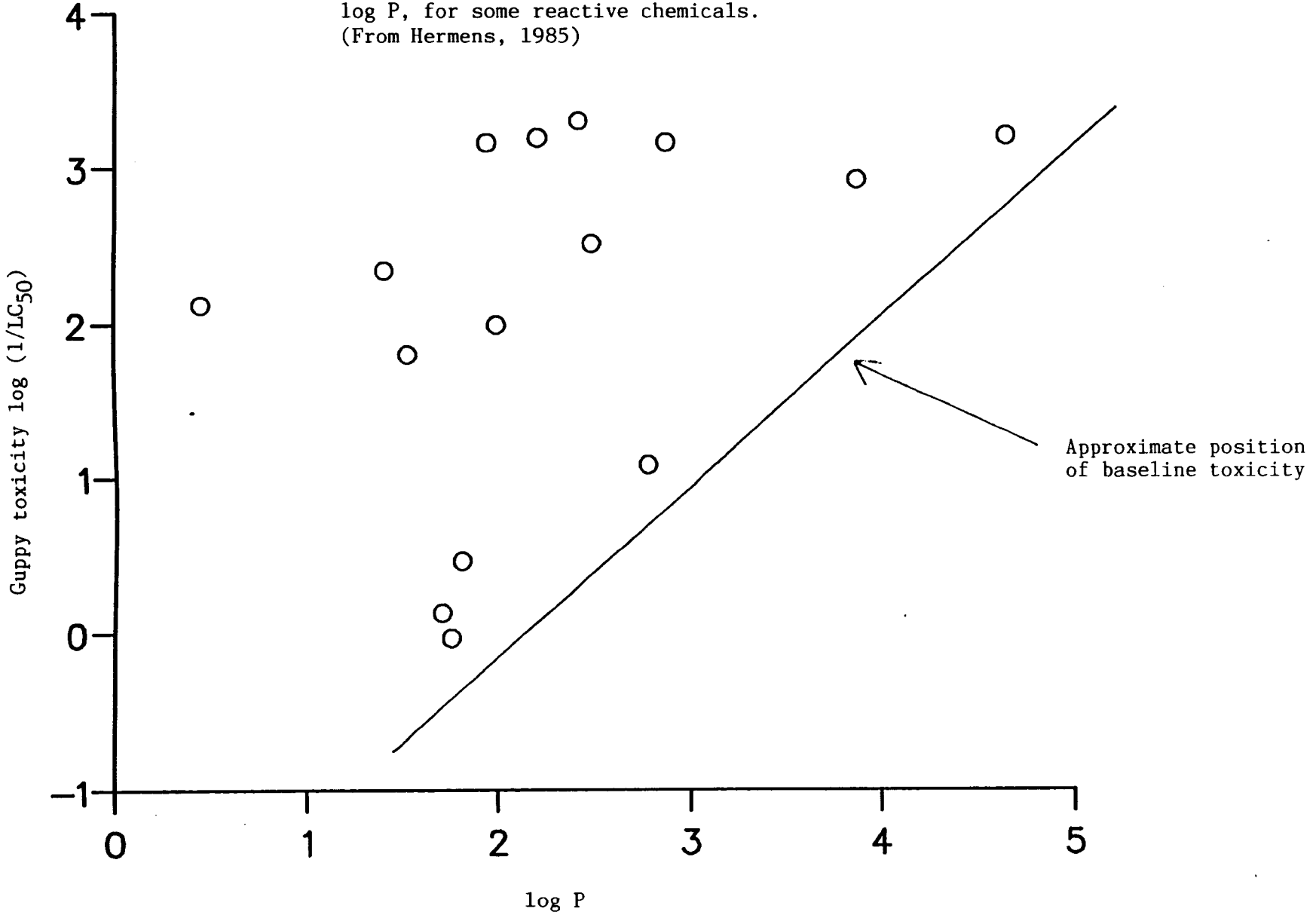
Where $LC_{50}(pred)$ is the predicted toxicity from log P
$LC_{50}(obs)$ is the observed toxicity

Chemicals with Te between 0.5-2 are considered to be acting by a narcosis mechanism, but with Te greater than 2, some form of more specific toxicity mechanism is likely.

1.6.3.2. Proelectrophile Toxicity Mechanism

Lipnick et al (1987) also proposed that for primary and secondary propargylic alocohols, which have been found to exhibit excess toxicity, the excess toxicity can be ascribed to a proelectrophile mechanism involving metabolism via the enzyme alcohol dehydrogenase to the corresponding alpha- and beta-unsaturated aldehydes and ketones which can act as Michael-type acceptor electrophilic

Figure 1.7 Graph of guppy toxicity against log P, for some reactive chemicals. (From Hermens, 1985)



toxicants.

When quantifying this an additional parameter is again required expressing the tendency of compounds to be reduced. Deneer et al (1987a) related the toxicity to fish of nitroaromatic compounds to hydrophobicity, the Hammett constant, and their polarographic half-wave reduction potential $E_{1/2}$. The best equation obtained was:

$$\log 1/LC_{50} = 0.96 \log P + 8.81 E_{1/2} + 0.68 \qquad (1.23)$$

n=20 s=0.18 r=0.964 F not given

1.6.3.3. Non-Polar Narcosis

Schultz and his co-workers have studied the effect of organic chemicals on the growth of the ciliate protozoan Tetrahymena pyriformis, with respect to investigating modes of toxic action (Schultz, 1988). As would be expected a strong relationship is found between the toxic effects of simple narcotic chemicals and the calculated log P from the MEDCHEM software (Schultz et al 1990b):

$$\log 1/IG_{50} = 0.834 \text{ ClogP} - 2.069 \qquad (1.24)$$

n=30 r^2=0.952 s=0.324 F=551

where IG_{50} is the concentration causing 50% inhibition in growth of Tetrahymena pyriformis after 48 hours

Other chemicals studied that were thought to be narcotics were, however, more toxic than estimated by this baseline narcosis model. These chemicals were observed to be more polar and had a hydrogen donor moiety (Veith and Broderius, 1987). Such chemicals are thought to act by a 'polar narcosis' mechanism and include para-, di- and tri- alkyl, and halogen-substituted phenols. Some phenolic derivatives (namely 4-nitro, 4-hydroxy, 2-hydroxy, 4-amino, and 2-amino) are not included because of their ability to form a Michael-type acceptor electrophile (Roberts 1987). In addition, Schultz et al (1989) have shown that alkyl- and halogen-substituted aniline derivatives can act by the polar narcosis method of toxicity.

A good overall relationship has been obtained for toxicity towards T. pyriformis

for 144 substitued phenols and anilines (Schultz et al, 1990b). It is suggested that the sigma constant reflects the fact that some of the phenols are partially ionised at the pH of the T. pyriformis growth assay. Since unionised molecules are better able to penetrate biological membranes than ionised species, this affects relative toxicity:

$$\log 1/IG_{50} = 0.641 \log P + 0.563 \Sigma\sigma - 1.19 \qquad (1.25)$$

n=144 s=0.240 r^2=0.891 F=571

where $\Sigma\sigma$ is the sum of the sigma constants of the substituents

1.6.3.4. Weak Acid Respiratory Uncouplers of Oxidative Phosphorylation

Veith and Broderius (1987) noted that some phenols were more toxic than predicted by polar narcosis. They postulated that many of these chemicals were acting as weak acid uncouplers of oxidative phosphorylation.

Respiratory uncouplers elicit their effect by abolishing the coupling of substrate oxidation to ATP synthesis. 2,4-dinitrophenol is thought of as a typical weak acid uncoupler, and other chemicals with a phenolic or anilinic moiety and additional electron withdrawing substituents can act in this manner. Such compounds include dinitro, tetra and pentahalogen-, phenylazo-, and dihalogen-mononitro- substituted phenols and anilines (Schultz et al, 1990b). Again a strong relationship is found for T. pyriformis toxicities:

$$\log 1/IG_{50} = 0.425 \log P + 0.202 \qquad (1.26)$$

n=26 s=0.176 r^2=0.912 F=249

Purdy (1988) showed the utility of parameters from computational chemistry to model nitrobenzenes suspected of acting as uncouplers of oxidative phosphorylation to the fathead minnow:

$$\log LC_{50} = -0.291 \log P + 0.569 E_{LUMO} + 5.37 C_{54} - 3.99 \qquad (1.27)$$

n=35 s=0.29 r^2=0.89 F not given

where LC_{50} is the 96 hour fathead minnow LC_{50}
E_{LUMO} is the CNDO calculated LUMO energy
C_{54} is the CNDO calculated lowest square of a nitro nitrogen P_z eigen vector for LUMO

1.6.3.5. Other Toxicity Models using Hydrophobicity

An interesting feature of these QSAR analyses is that many compounds, although not acting by a simple narcosis mechanism, can be well modelled by their hydrophobicity alone. This phenomenon has also been found by Deneer et al (1988) who found that the toxicity of aldehydes (thought to be chemically reactive *in vivo*) to the guppy was well modelled by log P:

$$\log 1/LC_{50} = 0.36(0.04) \log P - 2.54 \qquad (1.28)$$

n=14 s=0.19 r=0.923 F not given

They also observed that at higher log P values the compounds increasingly tend to act by a narcotic mechanism. This is because as log P increases there is a increasing tendency for accumulation in the lipid phase, thus becoming unavailable for interactions that occur in the aqueous phases of the organism. Cronin and Dearden (1990) also found that, with few exceptions, the toxicities of 17 different classes of organic chemical to the fathead minnow were well correlated to the calculated log P, despite several toxicity mechanisms being present.

1.6.4. Fish Acute Toxicity Syndromes (FATS)

Bradbury and his co-workers at the U.S. Environmental Protection Agency have taken a different approach to defining the mode of action of a toxicant using the assessment of fish acute toxicity syndromes (FATS). FATS (Bradbury, 1988) are distinct sets of rainbow trout (*Salmo gairdneri*) *in vivo* toxic responses that correspond to specific modes of action. By measuring a number of respiratory-cardiovascular variables, response sets associated with non-polar and polar narcotics, oxidative phosphorylation uncouplers, respiratory membrane irritants, respiratory inhibitors, acetylcholinesterase inhibitors, and central nervous system seizure agents have been created. The large number of reponses collated, such as oxygen consumption, ventilation, cough and heart frequency, has resulted in a large qualitative and quantitative data matrix. The use of discriminant function analysis has simplified this complex data set as well as enabling the best response variables for specific FATS to be determined.

1.6.5. Computerised QSAR Prediction

Commercially, there is only one real computerised toxicity prediction service available, TOPKAT, marketed by Health Designs Incorporated (HDI). TOPKAT is based on QSAR methodology (Enslein, 1988) and will (given a two dimensional structural input) provide estimates for the probability of the compound being carcinogenic, mutagenic, or teratogenic. Scores for eye and skin irritancy; and values for rat oral LD_{50}, fathead minnow LC_{50}, and Daphnia magna EC_{50} are also provided. For the prediction of continuous variables such as the rat oral LD_{50}, their approach can be considered as Hansch-type QSAR analysis in which standard parameters (ClogP, molecular connectivities, atomic charges) are combined with indicator variables for substructures or fragments. With a discrete property, such as the presence or absence of mutagenicity, discriminant analysis is applied.

Another method of predicitng discrete toxic responses is the computer automated structure evaluation (CASE) program devised by Klopman (1985). This selects automatically, substructural units that are most appropriate to discriminate between active and inactive molecules. In this system, the fragments are formed by breaking up the molecule into linear subunits consisting of three to twelve non-hydrogen atoms. Toxicity then can be assigned to specific fragments and structure searching can be used in a predictive mode.

Other computerised scoring systems based on QSAR have been proposed to help focus attention on existing chemicals needing urgent hazard assessment (Weiss et al, 1988), and to assess the environmental risk of new chemicals (Klein et al, 1988). This software has not yet, unfortunately, been made commercially available.

1.6.6. QSARs for the Prediction of Bioconcentration Factor (BCF)

In the study of the hazardous potential of xenobiotics to living organisms, the bioconcentration (or bioaccumulation) of the compounds is of particular importance. McCarty (1986) stressed the importance of bioconcentration as a factor in the prediction of aquatic toxicity and suggested that bioconcentration and toxicity kinetics are similar. The process of bioconcentration describes the

uptake of chemicals both from the ambient medium by direct contact and via food ingestion such that the resulting concentrations in the organisms are larger than those naturally occuring (Schuurmann and Klein, 1988). It is quantified as the ratio of the concentration of the chemical in the whole fish (C_f) to that in water at steady state (C_w) (Kenaga, 1972):

$$BCF = \frac{C_f}{C_w} \qquad (1.29)$$

It is well established that bioconcentration of chemicals in aquatic organisms is dependent on the hydrophobicity of the molecule, and bioconcentration has been well modelled by the partition coefficient (Metcalf et al, 1975). Examples of the correlation between hydrophobicity and bioconcentration are common. Neely et al (1974) found the bioconcentration factor for 8 chlorinated organic chemicals strongly log P dependent:

$$\log BCF = 0.542 \log P + 0.124 \qquad (1.30)$$

n=8 s=0.342 r=0.948 F not given

This model holds well for more complex organisms e.g. the bioconcentration potential of 8 organic chemicals in the adipose tissue of humans is well correlated to log P (Geyer et al, 1987):

$$\log BCF = 0.756(0.08) \log P - 1.415 \qquad (1.31)$$

n=8 s=0.261 r=0.969 F not given

Also Kerler and Schonherr (1988) found that for 7 lipophilic chemicals the bioconcentration in the cuticles of four plant species was well predicted by their octanol/water partition coefficient.

Some modifications have been made to the basic log P model. Koch (1983) reports the good correlation of the bioconcentration of 21 organic chemicals with first order valence corrected molecular connectivity. Also Anliker et al (1988) have found that the correlation is significantly improved with a $(\log P)^2$ and a steric term (in this case the molecular weight (MW) was used) for 43 organic dyes:

$$\log BCF = 0.82(1.26)\ \log P - 0.054(0.03)\ (\log P)^2 - 0.0048(0.001)\ MW + 0.88 \qquad (1.32)$$

n=42 s=1.01 r=0.65 F not given

Isnard and Lambert (1988) concluded that there is no significant advantage in using log P instead of aqueous solubilty. Schuurman and Klein (1988) found in a study of 49 diverse chemicals that log P is perhaps "overrated" and suggested solute accessible surface area and molar refractivity as better descriptors of bioconcentration.

1.6.7. QSARs for the Prediction of Biodegradability

Biodegradation (metabolism by microorganisms) is one of the most important processes determining the fate of organic chemicals in the environment (Alexander, 1981). Biodegradability rates, therefore, play an important role in the estimation of the environmental fate and hazard of chemicals. Biodegradation has been quantified by the calculation of psuedo first order, or second order, reaction rate constants for the degradation process. However, because of the difficulty in obtaining rate constants, the biological oxygen demand (BOD) for a sample of municipal sludge contaminated with the chemical has been utilised as a measure of microbial breakdown.

Existing QSARs for the prediction of biodegradability are reviewed by Parsons and Govers (1990). They report the success of QSAR in helping to determine mechanistic action. However, in general these apply only to restricted classes of compounds. For instance, Paris et al (1984) found the second order rate constant of biodegradability ($\log k_b$) for a series of alkyl esters of 2,4-D well correlated to hydrophobicity:

$$\log k_b = 0.799(0.098)\ \log P - 11.643 \qquad (1.33)$$

n=6 s not given r^2=0.972 F not given

Other descriptors model biodegradability well, including molecular connectivities (Boethling, 1986), e.g. for a series of phthalate esters:

$\log k_b = -37.2\ {}^2X + 547$ (1.34)

n=12 s=25.5 r=0.969 F not given

where ${}^2X$ is second order path molecular connectivity

Also Dearden and Nichoson (1987) report a very good correlation between the 5 day biological oxygen demand (BOD) of 197 compounds from 13 different classes and the difference in modulus of atomic charge across one specific bond ($[o]_{x-y}$) in a given chemical:

$\%\ BOD = 1.015 \times 10^3\ [o]_{x-y} + 1.523$ (1.35)

n=197 s=3.822 r=0.991 F not given

The value of QSAR to predict biological rates of degradation has, however, generally been limited to restricted ranges of chemical classes, and seems very dependent on the biological system utilised. Many different descriptors have been used to form relationships, varying from macroscopic physical properties to molecular structural parameters. For example, Pitter (1985) has found correlations with the Hammett constant, and Paris (1983) found biodegradation to be dependent on the substituent size of the phenols.

Reviewing the current literature, Govers and Parsons (1990) conclude that more information on the mechanisms, and rate determining steps of biodegradation, which can lead to a better founded choice of descriptors, and more biodegradation rate data, are required to further develop QSARs in this field.

1.7. Extrapolation of Toxicity Between Species

The extrapolation of animal toxicity data is a well established method to determine the risk of many products (e.g. consumer goods, drugs) to man. It aims to provide a safe upper concentration at which a chemical will cause no harm (Clayson, 1988). With the increasing need for as much toxicological information as possible, it was clearly appropriate to apply the approach to the field of environmental toxicology, where estimates are made in two major areas. The first is the area of effluent toxicity testing and monitoring, where an indication can be made as to the possible impact of a complex of chemicals to different organisms and ecosystems. The second is that of new chemical testing where there is a requirement to assess its potential hazard (Wallace and Niemi, 1988).
The second of these areas is of interest in this study. Much effort is being put into the extrapolation of biological activity in higher organisms from that in lower organisms (Tichy et al, 1985). These correlations of acute toxicities between unrelated species provide a means for estimating acute health hazards in species for which limited information exists. In addition, in comparing the relative sensitivity of species, this work has aided in the identification of the most suitable test species.
A summary of the main published work performed on inter-species correlations in environmental toxicology is shown in Table 1.3. The majority of the work has involved the prediction of quantitative values of toxicity, such as an LC_{50}, rather than attempts to find safe limits of chemicals as drug research tends to do. These relationships are normally expressed in the form of regression equations.
Correlation between toxicity to aquatic species have been relatively successful, although, as might be expected, there appears to be closer correlation between more closely related taxa. LeBlanc (1984), in an extensive study of inter-species correlations, found that when correlation indices derived from the comparison of sensitivity of a diverse group of aquatic organism to nonpesticide organic compounds were averaged into three categories:

Table 1.3 Summary of Inter-Species Correlations from the Literature

Organisms in the study *	Chemicals involved	Statistical methods applied	Reference
2, 9, 15	48 pesticides	model II regression analysis	Janardan et al 1984
1, 2, 11	32 inc. phenols, benzenes, anilines, solvents	linear regression	Hodson 1985
9-11, 15, 16, 24, 25	19 organic non-pesticides 30 pesticides 11 metals	polynomial regression	LeBlanc 1984
2, 8, 11, 25-28	7 organophosphorus insecticides	regression anlysis (including 1st order mol. connectivity)	Smissaert & Jansen 1984
11, 16	6 organic effluents from coal conversion process	ranking of toxicities	Black et al 1983
9, 11-13, 15, 17, 29-30	3 organic, 4 inorganics	ranking of toxicities	Holcombe 1983
9-10, 12, 14-15, 17-20, 25, 31	84 diverse organics	comparison of acute chronic (toxicity) ratios	Kenaga 1982
21, 32	57 phenols, anilines,	linear regression	Ahlers et al 1988
9, 25	6 surfactants, 1 detergent	linear regression	Maki 1979
1, 2, 11, 22,	11 organic pesticides	ranking of toxicities	Bathe et al 1976
11, (acute & chronic)	6 benzenes	linear regression	McCarty et al 1985
1-4	7 alcohols	linear regression (including 1st order mol. connectivity)	Tichy et al 1985
1-3, 5-7	18 potential anticancer drugs	linear regression	Freireich et al 1966
1, 2, 9, 11,	341 diverse organics	discriminant function analysis	Wallace & Niemi 1988
10-11, 13, 15, 23, 25, 33-34	33 phenols, 12 benzenes	linear regression	Ribo & Kaiser 1983
9, 33	31 priority pollutants	linear regression	Blum & Speece 1990
9, 33, 35	13 alcohols, 7 ketones	linear regression	Schultz et al 1990
9, 10, 25, 31	63 diverse organics and inorganics	errors-in-variables regression	Suter & Rosen 1988

* 1.mouse 2.rat 3.hamster 4.guinea pig 5.dog 6.monkey 7.man 8.pheasant 9.fathead minnow 10.sheepshead minnow 11.rainbow trout 12.brook trout 13.brown trout 14.lake trout 15.bluegill 16.largemouth bass 17.channel catfish 18.coho salmon 19.yellow perch 20.flagfish 21.golden orfe 22.crusian carp 23.guppy 24.algae (Selenastrum capricornutum) 25.Daphnia magna 26.stonefly 27.housefly 28.amphipod 29.snail (Aplexa hyporum) 30.chironomid (Tanyarsus dissimilis) 31.mysid shrimp (Mysidopsis bahia) 32.yeast (Saccharomyces cerevisiae) 33.microtox 34.bacteria (Bacillus sp.) 35.ciliate Tetrahymena pyriformis

i) closely related species (i.e. fish vs fish)

ii) more distantly related species (i.e. fish vs invertebrates)

iii) most distantly related species (i.e. fish vs algae)

the three averages were 0.82, 0.80, and 0.70 respectively.

There has, however, been some success in extrapolating from microbial organisms to higher taxa. Ribo and Kaiser (1983) concluded that the inhibition of light emission from photo-luminescent bacteria (the Microtox test) correlated well with other toxicity data for chlorophenols. Eqn 1.36 shows the relationship between 30 minute Microtox EC_{50} and the LC_{50} for semichronic toxicity to the guppy in a 7 to 14 day static test:

Guppy log LC_{50} = 0.79 log EC_{50} + 0.41 (1.36)

n=11 s not given r^2=0.89 F not given

Also Ahlers et al (1988) found for 15 phenols, anilines and aliphatics, a good correlation between golden orfe toxicity and the inhibition of the growth (IG_{50}) of the yeast Saccharomyces cerevisae:

Fish log LD_{50} = 1.25 log IG_{50} - 1.90 (1.37)

n=15 s not given r=0.94 F not given

Extrapolations of data between lower organisms and mammals are notoriously difficult (Tardiff and Rodricks, 1987), although some success has been achieved. Janardan et al (1984) found for 44 priority pollutants the following relationship between rat oral LD_{50} and bluegill 96 hour LC_{50}:

Rat oral log LD_{50} = 1.21 fish log LC_{50} + 0.539 (1.38)

n=44 s not given r=0.71 F not given

Of the authors listed in Table 1.3, only two have used regression techniques that account for possible errors in both variables (normal regression analysis accounts for errors in only the dependent variable). Janardan et al (1984) utilised Model II regression analysis; also Suter and Rosen (1988) employed an 'errors-in-variables' method. It may be preferable to use such a method as the tool to find inter-species correlations as it is certain that both sets of

toxicity data will contain errors.

Hodson (1985) found several correlations between rainbow trout oral and interperitoneal LD_{50} data and mouse and rat oral and interperitoneal LD_{50} data for some benzenes, phenols, and anilines:

Fish $IPLC_{50}$ = 1.00 Rat $IPLD_{50}$ - 1.69 (1.39)

n=16 s not given r=0.933 F not given

Extrapolation from fish to rat toxicity has also been improved by the use of first order molecular connectivity (Smissaert and Jansen, 1984; Tichy et al, 1985).

The use of structural descriptors in interspecies correlations has been taken a step further by Enslein et al (1987). They developed a structure-activity model to predict rat oral LD_{50} values from Daphnia magna LC_{50} values and various structural parameters including molecular connectivities and substructural keys. With 147 diverse organic chemicals the LC_{50} of D. magna and 12 structural parameters explained approximately 75% of the variance of the rat oral LD_{50}.

1.7.1. Relative Sensitivity of Species to Organic Chemicals

Analysis of comparative toxicity data can also give an indication of the relative 'sensitivity' of individual species to chemicals. A species is deemed to be more sensitive to a chemical than is another species if a lower concentration of the chemical produces a similar toxic response. Such species selectivity is, of course, an essential component in the design of many chemicals, e.g. pesticides have to be toxic to a particular species, yet should cause little harm to other species. Furthermore, marked differences in the susceptibility of species are commonly observed. Sloof et al (1983) report a factor of 9000 difference in the susceptibility of 22 aquatic species to 15 organic chemicals.

Other workers have assessed and quantified the relative senstivity of species. Holcombe et al (1987) reveal that the rainbow trout is the most susceptible of 12 species to 12 mixed organic chemicals and silver nitrate. Suter and Rosen (1988) examined the sensitivity of 21 species of marine fishes and crustaceans, finding that two shrimp species (Penaeus duorarum and Mysidopsis bahia) were on average

the most sensitive. Overall, however, no species was consistently found to be the most sensitive. This fact has led to the proposal that all new chemical testing should employ multi-species testing in preference to single species testing (Blanck, 1984; Cairns, 1988).

1.8 Aims of the Project and Areas of Investigation

The need for research into cheap, rapid, and reliable methods for the toxicological risk assessment of chemicals is crucial to help alleviate the workload on regulatory agencies worldwide. The thrust of this project has been to investigate two methods of achieving this goal, namely the use of QSAR, and the extrapolation of toxicity data between species. For both, high quality, accurate data are required for the biological activity that the chemicals exert and for the physico-chemical descriptors of each compound.

The Microtox test was chosen as a test system to investigate as it is a highly standardised, and reproducible procedure to assess aquatic toxicities of chemicals. The production of new toxicological data by this method complements other studies, as well as the regulatory data already available. It also has a proven track record in QSAR studies, and as a species from which data can be extrapolated. These data were augmented by a literature review of compatible toxicological methods.

The chemicals for analysis have to be very carefully chosen, especially in a QSAR study, in order to gain as much information from the data as possible. In addition to the investigation of narcotic chemicals, a particular emphasis in this study was put upon the modelling of those reactive organic compounds thought to be acting by more specific toxicity mechanisms. Obvious examples, for which limited toxicological information exists, are the nitriles, aldehydes, and the amines.

Little has been done to extend the techniques of computer chemistry and multivariate statistical analysis, founded in pharmaceutical and pesticide research to environmental toxicology. Use of molecular modelling can provide much valuable information, especially about such characteristics as the electronic configuration of reactive molecules; use of multivariate statistics can help to resolve the salient features of large quantities of data. All parameters used in the QSAR study were calculated, many of which had not been evaluated before in an environmental QSAR study. The use of calculated parameters has particular

advantages, such as they are cheap, relatively reliable, and can be rapidly produced for many compounds regardless of whether, or not, they have even been synthesised.

The extrapolation of toxicity data between species has been studied in this work, with a view to showing which species can give an accurate representation of the effects of the chemicals on higher species. Also the chemicals were assessed in an attempt to find which features, or chemical moieties, might induce a higher relative toxic reponse in one species than another. This aids the determination of the classes of chemicals for which the inter-species extrapolations can be applied.

2. METHODS

2.1. Determination of Biological Data

2.1.1. The Microtox Bioassay

A description of the background, theory, and application of the Microtox Bioassay is given in Appendix 1.

2.1.1.1. Summary of the Microtox System Principle of Operation

The Microtox System used the Microtox Model 2055 Toxicity Analyzer to measure the light output of the Microtox reagent. The Microtox reagent contains the living bioluminescent bacteria (Photobacterium phosphoreum) which have been grown under optimal conditions, harvested and then lyophilized (freeze-dried under vacuum). The lyphilized bacteria are rehydrated with the Microtox reconstitution solution to provide a ready-to-use cell suspension.

The Microtox Analyzer utilises a photomultiplier tube (PMT) to measure the light output of the luminescent bacteria before and after they are challenged by a sample of unknown toxicity. The Analyzer permits choice of test temperatures from 10°C to 25°C under ambient conditions.

The Microtox System is normally employed for the determination of a dose response curve, from which the effective concentration (EC) of a sample causing the specified effect is found. The basic procedure for this approach employs duplicates of a non-toxic control (the reagent blank) and four serial dilutions of the sample. The mean response of the duplicate reagent blanks is used to normalise the duplicate reponses of the four test concentrations of samples when the test results are reduced. This normalisation corrects the toxic response for normal drifts of light output with time and for small effects upon the light output arising from the dilution with the sample.

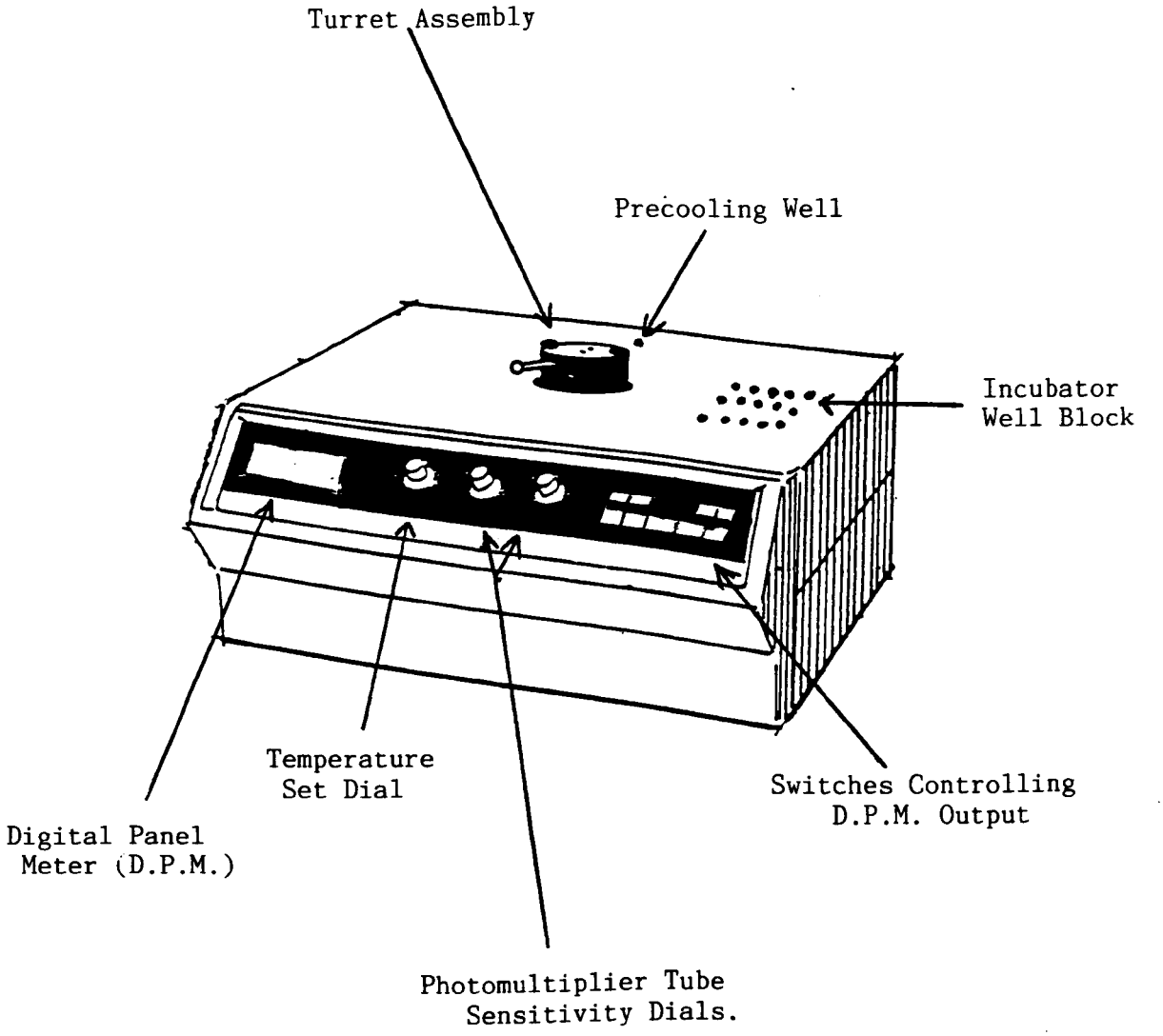
2.1.1.2. Features of the Analyzer

In this section the reader is referred to Fig 2.1.

Precooling Well

This incubator well is located behind the turret and is preset to 3 $\pm$1°C . It is used to precool the reconstitution solution and hold the reconstituted bacteria.

Figure 2.1 Diagrammatic representation of the Microtox Model 2055 Toxicity Analyzer.



The precooling well is situated inside the condenser which provides dry air for purging the turret and incubator wells of moisture. The temperature of the well and the condenser can be shown on the digital panel meter.

Incubator Well Block

The fifteen well incubator block is designed to maintain the cuvettes of reagent and sample at the selected temperature (in this case 15°C) prior to testing. The condenser provides dry air for purging the incubator wells. The most effective dry air purging is achieved when the incubator wells are empty or full of cuvettes.

Of the cuvettes in the fifteen wells, five are for the serial dilution of the sample (including a blank to act as a control), the remaining ten allow duplicate determination of the effect on light output of the sample.

Turret Assembly

The turret assembly is designed to accommodate one cuvette at a time for making a light reading. It contains an inner shutter which determines when the photomultiplier tube is exposed to the light from the cuvette. The inner shutter is only fully open when the turret is fully closed. The turret assembly contains a temperature controlled well.

Digital Panel Meter

The digital panel meter may be switched to display the relative light output, the high voltage level for the PMT, or the temperature of the precooling well, the turret well, or the incubator wells.

2.1.1.3. Operating Procedure

The Microtox bioassay was performed on the 48 chemicals listed in Table 2.1. There were many criteria for choice of the test chemicals. Most chemicals were chosen for which no previous Microtox toxicity data were available (only seven were repeats of published data). As well, it was felt important that there should be analysis of chemical classes for which no information exists, such as the aldehydes and nitriles. Also the compounds had to be considered as common

Table 2.1. Chemicals used in the Microtox Bioassay

Chemical	CAS no.	Stated Purity	% Methanol Added (if used)
3-pentanone	96-22-0	99%	
5-nonanone	502-56-7	98%	
3-methyl-2-butanone	563-80-4	99%	
3,3-dimethyl-2-butanone	75-97-8	95%	
2-ethoxyethyl acetate	111-15-9	99%	
methyl acetate	79-20-9	99%	
propyl acetate	109-60-4		
butyl acetate	123-86-4	99+%	
hexyl acetate	142-92-7	99%	
ethyl hexanoate	123-66-0	99+%	
diethyl adipate	141-28-6		
dibutyl adipate	105-99-7	96%	3
diethyl sebacate	110-40-7	95%	3
dimethyl malonate	108-59-8	97%	
diethyl benzyl malonate	607-81-8	97%	3
chloroacetonitrile	107-14-2	98+%	
malononitrile	109-77-3	99%	3
allyl cyanide	109-75-1		
1,4-dicyanobutane	111-69-3	99%	
1,6-dicyanohexane	629-40-3	99%	
octyl cyanide	2243-27-8	98%	3
acetone	67-64-1	99%	
toluene	108-88-3	99.5%	
2-methoxyethylamine	109-85-3	99+%	
1,2-diaminopropane	78-90-0	99+%	
butanal	123-72-8	99%	3
propylamine	107-10-8	98%	
2-chloro-4-methylaniline	615-65-6	98%	2
octylamine	111-86-4	97%	
hexanal	66-25-1	99%	
heptylamine	111-68-2	99+%	
4-fluoroaniline	371-40-4	99%	2
N,N-diethylaniline	91-66-7	99%	5
2-fluorobenzaldehyde	446-52-6	99%	
2-chloro-6-fluorobenzaldehyde	387-45-1	95%	3
5-bromosalicylaldehyde	1761-61-1	99%	3
vanillin	121-33-5	99%	5
2,4-dichlorobenzaldehyde	874-42-0		
4-chloro-3-nitrotoluene	89-60-1		2
1,2,4-trichlorobenzene	120-82-1	99+%	2
2-chloronitrobenzene	88-73-3	99+%	
3-chloronitrobenzene	121-73-3		2
2-chloro-4-nitrotoluene	121-86-8	98%	2
2-chloro-6-nitrotoluene	83-42-1	98%	2
acrolein	107-02-8	97%	
biphenyl	92-52-4	99%	5
1,3-dichloro-2-propanol	96-23-1	95%	
3-chlorotoluene	108-41-8		
4-chloronitrobenzene	100-00-5		2

Unless otherwise stated in this thesis, all undesignated alkyl chains are 1-(n-) substituted.

pollutants, and it was advantageous that comparative fathead minnow toxicity data were available.

All chemicals were supplied by Aldrich Chemical Co. Ltd, Gillingham, Dorset, U.K., except for 3-pentanone, toluene, and acetone supplied by BDH Co. Ltd, Poole, Dorset, U.K. The freeze-dried bacterial reagent was obtained from a local representative of Beckman Inc. The measurements of toxicity were taken at the Water Research Centre, Medmenham, Bucks, U.K.

Prior to administration the chemicals were diluted to form a stock solution using pure de-ionised water, and adjusted to 2% NaCl (with 20% NaCl solution) in order to prevent osmotic effects.

Increasing the Solubility of Chemicals

Due to their low solubility in water, 19 of the chemicals tested were prepared in an initial solution (i.e. before dilution) containing upto 5% methanol; the individual concentrations used are noted in Table 2.1. Solutions of 5-10% methanol have been shown to have no toxicity to the bacteria, and are recommended for use with compounds of low solubility (Ribo and Kaiser, 1987). Also results from this investigation show that 5% methanol has no effect on the relative light production of the reagent.

Analyzer and Sample Preparation

The Microtox Analyzer was set to 15^{o}C and allowed to equilibriate for at least one hour. 1.0 ml of Microtox reconstitution solution (a specially purified water) was placed in the cuvette in the precooling well. Cuvettes were placed in the incubation well block, and the Microtox Diluent (a specially purified water with 2% NaCl) added to the correct volume of the dilution factor. The osmotically adjusted primary dilution of the sample was then serially diluted in the appropriate cuvettes in the incubation well block.

Reagent Preparation

The Microtox Reagent was reconstituted. The precooled reconstitution solution was added to the vial of reagent (kept in cold storage for as long as possible), rapidly mixed, returned to the cuvette, and replaced in the precooling well as

not to allow warming of the mixture. The reagent was mixed again by aspirating and dispensing the contents of the cuvette with a 250 uL pipettor. 10 uL of reconstituted reagent was added to cuvettes in the incubator well block without sample present. All cuvettes were left for 15 minutes to stabilise.

Assay Procedure

The cuvettes containing the bacterial reagent were placed, in sequence, in the turret assembly so that a reading for the relative light output could be recorded on the digital panel meter. Immediately after the last reading was taken, 500 uL (or the necessary volume according to the dilution factor) of the appropriate sample dilution (or blank) was added and mixed with the appropriate cuvette of bacteria. This operation was undertaken in the same sequence, and at the same time intervals (fifteen seconds) as the light output readings. Five and fifteen minutes after the first sample had been added, the light output was again recorded in the same sequence. This gives an indication of the effect (if any) of the sample at each time.

Experimental Design

Each chemical was given a 'preliminary test', using a large range of dilutions, e.g. a 10:1 dilution scheme, to establish a range of concentrations spanning the EC_{50} of the toxicant. On this was based a more accurate investigation, using a more specific range of concentrations (normally a 2:1 or occasionally 5:1 dilution scheme), to establish an accurate EC_{50} of the compound.

The preliminary tests were performed at least twice, further repeats being carried out if the results were over 25% divergent. The more accurate tests were performed three or more times, with at least two different initial solutions used in order that any errors in the preparation of the solution could be identified.

2.1.1.4. Analysis of the Data

For each analysis ten data were produced. These were reduced to a express the effective concentration at which there is a 50% decrease in the relative light

intensity (EC_{50}). To account for the natural decay in the output of light from the bacteria (due to time etc.), the readings were normalised using a 'blank ratio' (BR), which is the ratio of actual reading for the control solutions without toxicant at the start of the assay and at each time a recording is made:

$$BR_t = \frac{I(0)t}{I(0)o}$$

Where I(0)o, I(0)t are the blank readings at time 0 and t.

The gamma function, the ratio of light lost due to the effect of the sample as compared with the light remaining from the blank can then be calculated and used as a bioassay response parameter. Gamma is calculated by:

$$\text{gamma} = \frac{(BR \;.\; I(c)o) - I(c)t}{I(c)t}$$

Where $I(c)_t$ is the intensity of light measured for a cuvette containing sample concentration c after t minutes (c=0 indicates the blank).

When the gamma function is plotted against the logarithm of the corresponding concentration of the toxicant, the EC_{50} is found when gamma = 1.

A worked example is given below for the analysis of the 5 minute EC_{50} for a 10:1 dilution of N,N,-diethylaniline. The full results are given in Table 2.2.

Thus the blank ratio after 5 minutes is the mean of the duplicates:

For 1. BR = 77/87=0.88

For 2. BR = 101/114=0.88

The mean of the blank ratio is 0.88.

For the first duplicate of the 0.00089 ml/L concentration gamma is calculated by:

$$\text{gamma} = \frac{(0.88 \;.\; 126) - 75}{75} = 0.47$$

and so on for the other concentrations.

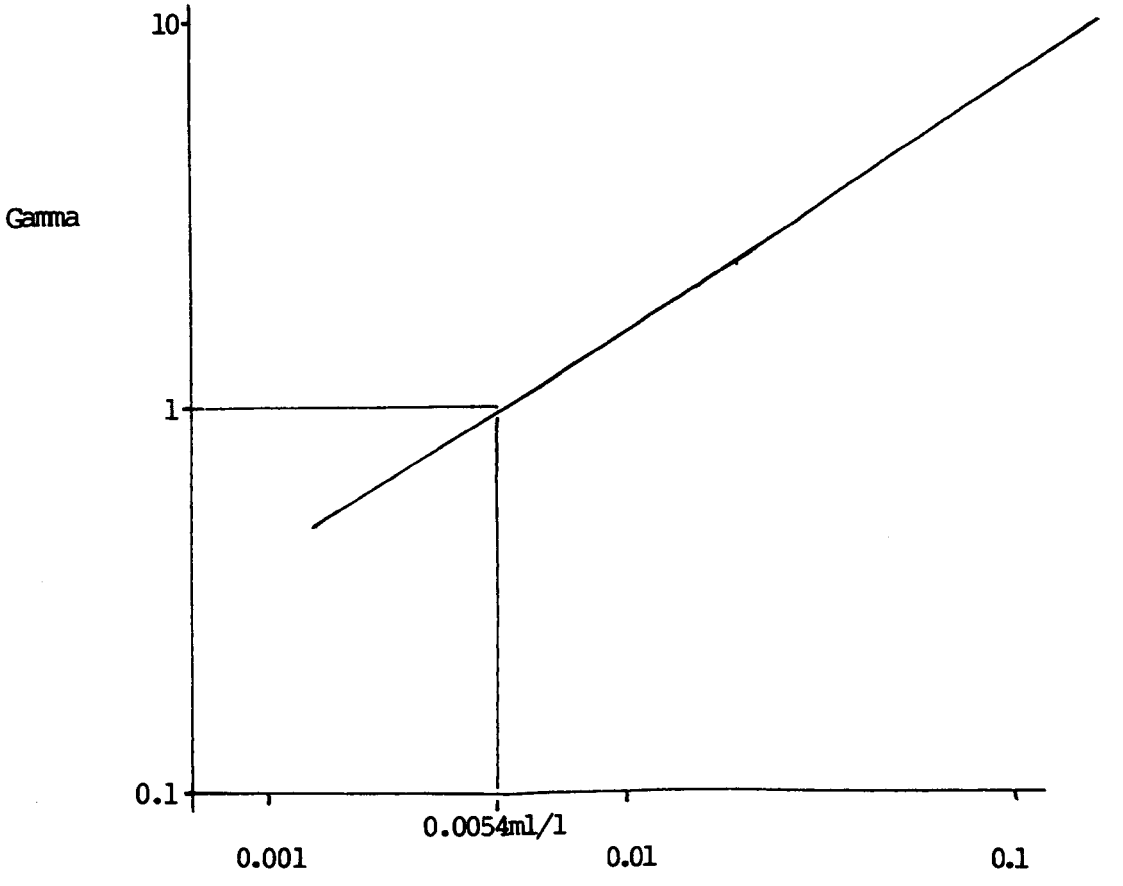
Table 2.2. Full Results of the Light Output of Bacteria when Subjected to Different Concentrations of N,N-diethylaniline

Concentration of N,N-diethylaniline		Light Output (from the digital panel meter) Initial	5 minutes	Gamma
Blank	1	87	77	-
	2	114	101	-
0.00089 ml/L	1	126	75	0.47
	2	98	69	0.26
0.0089 ml/L	1	101	36	1.48
	2	108	42	1.27
0.089 ml/L	1	83	6	11
	2	105	9	9
0.89 ml/L	1	103	4	22
	2	114	4	24

The plot of gamma values against the logarithm of the concentration of N,N,-diethylaniline is shown in Fig 2.2. The line is plotted through the mean of the gamma values and shows the EC_{50} where gamma = 1 to be at a concentration of 0.054 ml/L.

Fig 2.2. Plot of the gamma function of light output for a 10:1 dilution of N,N-diethylaniline

N.B. Both axes are in log units.



2.1.2.1. Calculation of Fathead Minnow 96 hour LC_{50}

The problem of obtaining toxicity data of a reliable quality and quantity is considerable (Cairns, 1988). However in a far-sighted move in the 1970's administrators at the Office of Toxic Substances in the U.S. Environmental Protection Agency (E.P.A.) realised the need for a reproducible, accurate toxicity data base. In 1979 the University of Wisconsin-Superior embarked on a data generation effort to fulfill this major need in toxicology. The primary toxicity endpoint selected was the 96 hour lethal concentration LC_{50} for the fathead minnow (Pimephales promelas). The fathead minnow was chosen as the test species due to its ease of culture, widespread occurrence, rapid growth, ecological importance, and mid-range in tolerance for freshwater organisms to environmental pollutants (Geiger et al, 1988). In total the toxicities of over 550 organic chemicals had been measured by 1988 when the project ceased. These data are made freely available for research purposes, including QSAR studies such as this.

A brief overview of the method is given below (for full methods the reader is referred to Brooke et al, 1984; Geiger et al 1985, 1986, 1988). The acute toxicity tests were conducted according to ASTM recommendations (ASTM, 1980). Juvenile fathead minnows ranging in age from 29 to 33 days were fed live brine shrimp nauplii in excess until 24 hours prior to testing, and were not fed during the 96 hour exposure. Tests were performed with a continuous-flow diluter exposure system. This comprised of four replicate glass exposure chambers, through which the relevant concentration of the toxicant was passed. (Benoit et al 1982). Each toxicity test had five treatment levels and one control, twenty fish being placed in each tank.

Lake Superior was the source of dilution water. The water was filtered through sand and heated to $25\pm2^{o}C$ for the tests. The water chemistry was analysed, and the mean pH, dissolved oxygen, temperature, alkalinity, and hardness were recorded. The compounds used in the study were obtained from several chemicals suppliers, and were normally of the purest available form, these purities were

confirmed by gas chromatography.

All LC_{50}s and 95% confidence limits were calculated using the average of the analysed tank concentrations which were corrected for recovery, and by the computerised Trimmed Spearman-Karber method for estimating median lethal concentration.

During each test the fish were observed at 8, 24, 48, 72, and 96 hours for 98 parameters of abnormal behaviour and morphological changes. The behavioural data were statistically examined and classified into possible modes of fish acute toxicity syndromes - FATS (Drummond et al, 1986).

2.2 Calculation of Physico-Chemical Data

2.2.1. Calculation of the Partition Coefficient and Molar Refractivity

2.2.1.1. Software

The suite of programs incoroprated into the MEDCHEM software has been pioneered by Leo and his co-workers at the Pomona College, California. MEDCHEM has become, in effect, the standard method for calculating the logarithm of the 1-octanol/water partition coefficient and molar refractivity, being extensively used in academia and the pharmaceutical industry worldwide.

MEDCHEM (ver 3.53) was run interactively on a Microcolour M2250 graphics terminal (emulating a VT241) supported by a VAX mainframe.

2.2.1.2. Calculation of the Logarithm of the Octanol/Water Partition Coefficient (ClogP)

ClogP is calculated using the 'constructionalist approach' pioneered by Leo et al (1975). This starts with a set of 'fundamental fragments' whose values are summed with appropiate weighting factors consisting of the number of times each fragment occurs. Correction factors are added as necessary. This method can be described by:

$$\log P = \sum_{i=1}^{n} a_i f_i + \sum_{j=1}^{m} c_j$$

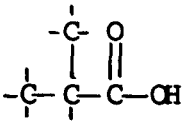
where f_i is the fragment constant for the i^{th} fragment
a_i is the number of occurances of the i^{th} fragment
c_j is the j^{th} correction factor

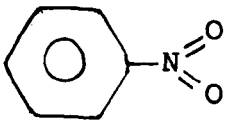
The fundamental assumption of the 'constructionalist approach' is that hydrophobicity is an additive-constitutive property of molecules. Thus the log P value is equal to the summation of the hydrophobic contributions of each constitutive fragment. The MEDCHEM software contains fragment constant values, and correction factors which it applies to each substructural fragment as appropriate.

2.2.1.3. Input of a Structure

Structural input of compounds is very flexible, methods include a simple name, or identification numbers such as CAS or Aldrich ID. However, when these are not

available, or not found in the data base, a simple structural input is employed, known as SMILES. SMILES (Simplified Molecular Input Line Entry System) is a chemical notation system designed for modern chemical information processing (Weininger et al, 1986). It encodes the molecule in the form of a simple hydrogen- suppressed string. Each atom is represented by its atomic symbol. Other symbols used include = for a double bond and # for a triple bond. Lower case implies aromaticity, and branches are shown in parentheses. Cyclic structures are represented by breaking a bond in each ring structure. The bonds can be numbered in any order, designating ring opening (or ring closure) bonds by a digit immediately following the atomic symbol at each ring closure. The molecule can be written from any starting point, meaning many different but equally valid descriptions of the same structure can exist, as long as the connectivity of the molecule is maintained. Unique coding is thus possible for many molecules, including structural isomers. However until isometric SMILES is available, it cannot differentiate between stereo (i.e. cis-trans and optical) isomers.
Two examples of SMILES notation are given below:





Name: Isobutyric acid Nitrobenzene

SMILES: CC(C)C(=O)O c1ccccc1(N(=O)=O)

After running the program, the output incorporates a graphical representation of the chosen molecule; its name or SMILES coding; a summary of the information contained in the THOR data base (if available); and calculated estimates of the partition coefficient and molar refractivity. THOR, ClogP, and CMR are described below. If, however, the appropiate fragment constants are not found in the data base, log P cannot be calculated.
A sample of output is shown in Table 2.3, which shows the constants applied to

Table 2.3. Sample of MEDCHEM Output Demonstrating the Fragmental Constant Approach for Isobutyric Acid

Number of Fragments and their Description	Hydrophobic Fragment Value
1 Carboxy group	-1.110
3 Aliphatic isolating carbons	0.585
1 Non-halogen, polar group branch	-0.220
7 Hydrogens on isolating carbons	1.589
2 Chains (bonds)	-0.240
Total value for ClogP	0.604

each fragment, their sum being the estimate of the partition coefficient. The first four fragments are self-explanatory, the fifth, referred to as "chains (bonds)" is a value for the number of bonds made by a non-ring carbon atom that are not doubly, or triply-bonded to a heteroatom.

2.2.1.4. THOR (THesaurus Oriented Retreival)

Incorporated into the MEDCHEM software is the ability to access a large chemical information database (THOR). THOR is a database specifically designed for efficient storage and retrieval of chemical information based on a compound's structure. When a compound is 'inputted' into the software the database is searched for a match. If the compound is found, it is possible to view information such as CAS number; local, common and trade names, and measures of the partition coefficient (including those in solvents other than octanol). THOR displays one of the measured log P values (that considered to be most accurate) alongside the calculated value.

2.2.1.5 Calculation of Molar Refractivity

Molar Refractivity (MR) can be considered as the sum of either atom or bond refractivities. It is calculated from the Lorentz-Lorenz equation:

$$MR = \frac{u^2 - 1}{u^2 + 2} + \frac{MW}{d}$$

where u is the index of refraction
MW is the molecular weight
d is the density

Although the measurement of u and d is simple, they cannot, of course, be obtained if the compound has not yet been synthesised, or is in short supply. Hence there is a need for a calculated value of MR. Basically, fragment based constants are applied to a structure, in the same manner as ClogP is calculated, and the sum, together with correction factors for some bond types, is assumed to be an estimate of molar refractivity.

The origin of this method was a linear regression analysis performed on a 'training set' of the molar refractivity (calculated from the Lorentz-Lorenz equation) of 1400 compounds. The analysis investigated the effect of various indicator variables (that characterise the structure) on molar refractivity, and provided an estimate for the molar refractivity of each fragment (Medchem Software Manual, 1987).

2.2.2.1. Molecular Connectivities

Molecular connectivities and Kappa shape indices were calculated for the compounds using the MOLCONN2 (ver 1.0) program running on a VAX mainframe. This is an interactive program supplied by Hall Associates Consulting, Quincy, MA. A brief overview of the calculation of molecular connectivities and Kappa indices is given in this section. For a more complete description of their calculation, application, and nature see Appendix 2.

2.2.2.2. Calculation of Molecular Connectivities

Molecular connectivities are calculated from a knowledge of the atom connections in a molecule. Their calculation is explained briefly below, and in more detail elsewhere (Kier and Hall, 1986).

First order connectivity (^1X) is calculated thus

$$^1X = \sum (d_i \; d_j)^{-0.5}$$

where d is the number of non-hydrogen bonds of atoms i and j, when considered across one bond

The sum of the function $(d_i \; d_j)^{-0.5}$ is termed the first order molecular connectivity. Higher orders are calculated across 2 bonds (for second order), 3 bonds (for third order) etc. Also considered are atoms connectivities across features such as clusters, and path/clusters.

A second class of molecular connectivity (termed valence corrected) is calculated in a similar manner, except that the d (delta) value counts all bonds (e.g. a count of two for a double bond, three for a triple bond) made to atoms other than hydrogen.

2.2.2.3. Calculation of Kappa Indices

The Kappa indices attempt to parameterise the shape of a molecule. They are a consideration of the number of atoms in the molecule (A), and the number of paths of length one (^1P) for first order, length two (^2P) for second order, and length three (^3P) for third order in the hydrogen suppressed graph of the skeleton structure. (Kier, 1985; 1987)

Thus the Kappa values are given by

$$^1K = A(A-1) / (^1P)^2$$

$$^2K = (A-1)(A-2)^2 / (^2P)^2$$

$$^3K = (A-3)(A-2)^2 / (^3P)^2 \qquad A \text{ is even}$$

$$^3K = (A-1)(A-3)^2 / (^3P)^2 \qquad A \text{ is odd}$$

$Kappa_\alpha$ (alpha) values encode a modification to account for non-hydrogen atoms other than Csp^3. This involves recalculating A by the value α α being defined as the ratio of covalent radii of atom x relative to Csp^3, or

$$\alpha = r_x / r(Csp^3)-1$$

The $Kappa_\alpha$ values are recalculated using this term.

2.2.2.4. Input of a structure

A compound is entered into the MOLCONN2 program as coding written in an edit file, known as the B (or bond) file (later versions will use SMILES coding). The coding is quite simple; each non-hydrogen atom is numbered, and the number of hydrogens attached, its atomic symbol, and the identity numbers of other non-hydrogen atoms to which it is connected are listed. Two more lines of coding are needed, the first line gives an identification number and title, the last line, simply -1, terminates the input. The numbering can start at any atom in the molecule, although experience has shown it is better to number the basic structure first, followed by any branching or substitution. The program will differentiate between structural isomers, however not between stereo isomers. Two examples of B files are given in Table 2.4.

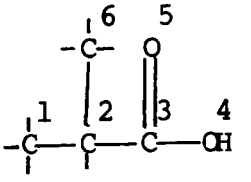
2.2.2.5 Output of Data

The results of the calculations are given in two forms, a listing file, and a file for statistical analysis. The listing file gives full explanation of the results. The file for statistical analysis is in a standard form, without text, summarising the results. With a small amount of manipulation, the statistical analysis file can be entered straight into the statistical routine required, thus eliminating tedious manual data entry, and errors that may occur from it.

Table 2.4. Examples of the B File for an Input into the MOLCONN2 Program

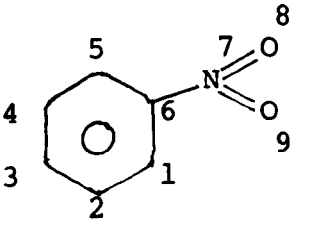
1. Isobutyric acid

```
1,ISOBUTYRICACID
1,3,C,2
2,1,C,1,3,6
3,0,C,2,4,5
4,1,O,3
5,0,O,3
6,3,C,2
-1
```



2. Nitrobenzene

```
2,NITROBENZENE
1,1,C,2,6
2,1,C,1,3
3,1,C,2,4
4,1,C,3,5
5,1,C,4,6
6,0,C,1,5,7
7,0,N,6,8,9
8,0,O,7
9,0,O,7
-1
```



The non-hydrogen atoms of each molecule are numbered according to the B file.

2.2.3 Chemical Descriptors from Computational Chemistry

2.2.3.1 Molecular Modelling Software

Molecular modelling and the calculation of chemical descriptors were performed using the COSMIC (COmputational Structure Manipulation In Chemistry) computer package. COSMIC (Vinter et al, 1987) is a complete integrated framework of molecular modelling and computational chemistry software. It was originally devised by Dr A.J. Vinter and his co-workers as the 'in-house' modelling system at the Wellcome Foundation in 1976. The program is being continually updated and has benefitted from the contributions of Drs A. Davies and M. Saunders of Smith, Kline, and French Research Ltd. More recently it has been distributed, free of charge, to academic institutions under a strict licensing agreement, the release controlled by Dr D. Jackson, Department of Pharmacy, Nottingham University. COSMIC is a suite of programs that allows molecular modelling and analysis from first principles. After a chemical structure is inputted, its geometry can be optimised using either molecular mechanics, or molecular orbital theory methods. Various molecular orbital methods are available which also enable partial atomic charges etc. on the molecule to be calculated. As well, highly developed display routines allow the molecule to be observed, and different steric and electronic functions to be assessed.

Some of the software, such as the structure entry environment and graphics display routines were specifically written for the system by the authors. Other programs, such as the molecular orbital methods, are those supplied by the Quantum Chemical Program Exchange (QCPE), Chemistry Department, Indiana University, for a nominal charge, which have been adapted to interface with the system.

2.2.3.2. Input of a Chemical Structure

COSMIC was run on a Microcolour M2250 graphics terminal (emulating a TEK 4105) supported by a VAX mainframe. Compounds were constructed graphically into the DRAW environment of COSMIC using a mouse. Aliphatic compounds were entered as a non-hydrogen skeleton structure. Aromatic compounds were created in a similar

fashion; the benzene structure was taken from the COSMIC fragment library and non-hydrogen substituents were added. The hydrogen-suppressed skeleton structures were roughly minimised with a first derivative minimiser (utilising a Newton-Raphson/ Simplex method) for 99 iterations or until the root mean square of the energy converged within 1.00 cal/mole. Hydrogens were added to the skeleton and the geometry of the compound was optimised more precisely by the MIN02 quasi Newton-Raphson method (White, 1977). Partial atomic charges were calculated for the compounds by two methods. CNDO (Complete Neglect of Differential Overlap) calculations were performed on the CNDO/2 routine written by Pople and Beveridge (1970), which has been largely recoded to run faster. Standard settings of charge, multiplicity and convergence criterion were chosen. MNDO (Marginal Neglect of Differential Overlap) calculations of partial atomic charge were also obtained from the AMPAC routine (Dewar Research Group, 1986); MOPAC keywords ISCF and ENPART were employed. After having the partial atomic charges calculated, the molecules were re-minimised using the MIN02 routine.

2.2.3.3. Calculation of Parameters

It is common in the study of a homologous series of compounds, such as is often available in pharmaceutical (De Benedetti, 1987) or pesticide research (Ford et al, 1989), to observe and attempt to model the effect of different substituents on the charges, superdelocalisabilities and other properties of the basic structure. In this project, however, because of the diverse nature of the chemicals it was not possible to utilise the individual atomic data. Whole molecule parameters were therefore obtained using the 'MO Options' routine in COSMIC that displays eigenvectors, dipoles, and Huckel reactivity indices. These parameters included HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital) energies and their difference (HOMO and LUMO are thought to approximate the electron donating and accepting capabilities of the molecule respectively); dipole moment; the total electronic energy of the minimised molecule; and the polarisability calculated from the sum of the self-

atom polarisabilities.

The capability of obtaining the 'theoretical optimum structure' of a molecule gives an ideal opportunity to calculate various measures of the size, or bulk of a compound. The accessible surface area of the molecule was obtained by the method of Pearlman (1981). Here the atoms of the molecule are considered as intersecting spheres, and the area of the exposed surface is calculated. Overall molecular volume and dimensions (Edward, 1970) obtained included the van der Waals volume, alternative molecular volume, collision diameter, and diameter of closest approach. Finally Monte Carlo areas and volumes (Smith, 1986) were computed, with a probe radius of 0.0. All these calculations were performed in the 'Atom Manipulation' environment of COSMIC.

2.3. Quantitative Structure-Activity Relationships for Fathead Minnow and Microtox Toxicities

2.3.1. The Data Matrix

There are 49 descriptors for each chemical (see Table 2.5) including measures of hydrophobicity - ClogP, size - molecular connectivities, Kappa values, and COSMIC steric values, and electronic terms - HOMOs, LUMOs, Dipole Moments. However, using this number of variables brings with it inherent problems for statistical analysis. The dangers of using a relatively 'over-square' data matrix are well chronicled (e.g. Kikuchi, 1987), and include the possibility of chance correlations. It is often possible to reduce the dimensionality and collinearity of such a data set, without significant loss in the information content. The approaches taken and their relative performance to achieve this are described below.

Many (mainly multivariate) statistical methods are available for analysing such a data matrix (see Section 1.4.3), either by data reduction or preparation of new uncorrelated variables. Livingstone (1989) identified the principal statistical techniques employed for this purpose (see Table 1.1). Due to software limitations and the type of data available only cluster analysis on the variables, principal component analysis, factor analysis, canonical correlation, and multiple linear regression were attempted. Methods 2, 3, 6 and 10 in Table 1.1 are known as pattern recognition techniques and are suited to determine 'categoric', as opposed to linear, data. Also these require specialist software not yet available at Liverpool Polytechnic.

2.3.1.1. Cluster Analysis

Cluster analysis on the variables is a good method of reducing their dimensionality and collinearity, while helping to increase understanding of their structure. Initially each variable is considered as a separate cluster, then the two most similar variables are joined to form a cluster. The amalgamating process continues in a stepwise fashion (joining variables or clusters of variables) until a single cluster is formed that contains all the variables. The measure of

Table 2.5. Summary of Chemical Descriptors used in QSAR Analysis and their Abbreviations used in this Study

Steric

zero, first, second and third order path molecular connectivity (PS0, PS1, PS2, PS3)
third order cluster molecular connectivity (CS3)
zero, first, second and third order valence-corrected path molecular connectivity (PV0, PV1, PV2, PV3)
third order valence-corrected cluster molecular connectivity (CV3)
difference between simple and valence path molecular connectivity for each of zero, first, second and third orders (P(S-V)0, P(S-V)1, P(S-V)2, P(S-V)3)
difference between third order cluster simple and valence molecular connectivity (C(S-)3)
sum of simple and valence path molecular connectivity for each of zero, first, second and third orders (P(S+V)0, P(S+V)1, P(S+V)2, P(S+V)3)
sum of third order cluster simple and valence molecular connectivity (C(S+V)3)
first, second and third order Kappa value (K1, K2, K3)
zero, first, second and third order Kappa alpha value (Ka0, Ka1, Ka2, Ka3)
calculated molar refractivity (CMR)
Accessible Surface Area (ASA)
van de Waals volume (VWvol)
alternative molecular volume (AltMV)
collision diameter (CollDia)
closest approach (ClosApp)
Monte Carlo area (Area)
Monte Carlo volume (Volume)
molecular weight (MW)

Electronic

CNDO calculated dipole moment (Dipole)
CNDO calculated total electronic energy (Energy)
CNDO calculated LUMO (LUMO)
CNDO calculated HOMO (HOMO)
difference in CNDO calculated HOMO and LUMO (DiffH-L)
CNDO calculated whole molecule polarisability (Polariz)
MNDO calculated dipole moment (MDipole)
MNDO calculated total electronic energy (MEnergy)
MNDO calculated LUMO (MLUMO)
MNDO calculated HOMO (MHOMO)
difference in MNDO calculated HOMO and LUMO (MDiffH-L)
MNDO calculated whole molecule polarisability (MPolariz)

Hydrophobic

calculated logarithm of 1-octanol/water partition coefficient (ClogP)

The letters in parentheses are the abbreviations by which the descriptors are commonly refered to throughout this thesis.

similarity is the absolute value of the correlation and the clusters were joined using the minimum distance rule (single linkage). The analysis was run using program P1M of the BMDP statistical software available at the University of Manchester Regional Computer Centre.

Cluster analysis was performed twice, with the data corresponding to both toxicological endpoints. Clusters were formed and analysed at an arbitrarily chosen similarity level of 90%. This level of similarity allows a great reduction in the amount of data, yet will hopefully maintain sufficient information for meaningful regression analysis to be applied. One of the variables from each cluster was then chosen to represent that cluster in a 'decorrelated' data set. The choice of the variable was based on experience of which was likely to be the most reliable, and useful, variable. The 'decorrelated' data were then available to be put into the stepwise regression analysis.

2.3.2. Stepwise Regression Analysis

Forward stepwise regression analysis was performed with the Microtox and fathead minnow toxicities as dependent variables, and the decorrelated chemical descriptors from cluster analysis (see Tables 3.3 and 3.4) as independent variables. The analysis was run using the MINITAB statistical software (ver 7.1) with a F-to-enter the equation for each variable of 4.0. Each analysis was confirmed using best-subsets regression analysis. Stepwise regression analysis was also performed on the toxicities according to the classes to which they were assigned (see Table 3.2). Again the decorrelated chemical descriptors were utilised.

2.3.3. Principal Component Analysis

When collinearity is present in a data set such as that obtained in this study, it is profitable to obtain a reduced number of new uncorrelated variables, that contain as much of the original information as possible. Principal component analysis and factor analysis are two methods of achieving this objective. Principal component analysis (PCA) is often used to give a preliminary understanding of the data set. The correlation matrix of the variables is reduced mathematically to a unique set of eigenvalues, or principal components. These have several important properties detailed below.

i) each eigenvalue is made up of a linear combination of the original variables.

ii) these eigenvalues are orthogonal to each other in multidimensional space, representing different dimensions in it, and thus are totally uncorrelated.

iii) there are as many eigenvalues of a matrix as there are rows and columns in the matrix.

iv) the sum of all eigenvalues is equal to the sum of the diagonal elements of the matrix. Since these are unity, the sum of the eigenvalues is equal to the number of variables.

v) the eigenvalues will, however, give no information on which data are redundant or irrelevant.

The importance of each eigenvalue is expressed as the fraction of the total variance of which it explains. This is defined as

$$FV_{i-j} = \frac{\text{Eigenvalue}_i}{\sum \text{Eigenvalues}_{i-j}}$$

where FV_i is the fraction of the variance explained by eigenvalue i
$\sum$ Eigenvalues$_{i-j}$ is the sum of all eigenvalues

The eigenvalues are created such that the first explains as much of the variance as possible, the second the second largest amount of variance, etc. A general 'rule of thumb' is that a significant principal component will have an eigenvalue of greater than one (Martin, 1978), since a principal component with an eigenvalue of less than one 'explains' less variation in the original data than

does one of the test scores (Manly, 1986).

The major drawback of using PCA in this context is, however, the loss in descriptive power of each newly formed variable. As a combination of all the variables, it no longer represents a single feature of the molecule, as an individual variable does. Instead information must be obtained from the 'loadings' or eigenvectors of the original variables onto the principal components, the relative size of which will give an indication of the properties of that eigenvalue. Thus a principal component with large eigenvectors for steric variables can be assumed to describe the size, or bulk, of the chemicals, and so on.

Results from these principal components (known as scores) for each chemical can be utilised as new chemical descriptors in a number of ways. The scores are used in multiple regression analysis, against scalar properties such as toxicity, or if plotted against each other provide a basis for pattern recognition of categoric data such as carcinogenicity.

2.3.3.1. Principal Component Analysis in the QSAR Study

Principal component analysis was performed using the program P4R in the BMDP statistical software. This involved separate principal component analyses being carried out on the 49 physico-chemical descriptors (summarised in Table 2.5) associated with each of the fathead minnow and Microtox data sets. Best-subsets regression analysis was then run for toxicities against the scores of the principal components with an eigenvalue greater than one (on MINITAB ver 7.1).

2.3.4. Factor Analysis

The basic rationale of factor analysis (FA) is similar to principal component analysis in that it sets out to produce novel orthogonal variables (factors) from a highly collinear data matrix. However, it differs from principal component analysis in some fundamental ways concerning the formation of the factors. Factor analysis attempts to explain the co-variance between the variables themselves, whereas principal component analysis explains the total variance of the data. Also, unlike PCA, FA is based on a proper statistical model (see Manly (1986) for full derivation) which assumes that the total variance of the data can be accounted for as the sum of the variance common to all variables and that unique to each variable. Thus factor analysis is concerned only with describing the 'common' variance.

FA conforms to the following criteria. The principal components are calculated for the data matrix up to a specified level of significance. (In common with PCA this is normally assumed to be an eigenvalue of greater than one). The eigenvectors of the principal components are then transformed into the corresponding unrotated factor pattern by multiplying the vectors by the square root of the corresponding eigenvalue. These factors are then 'rotated' to produce a rotated factor matrix in which there are more high and low coefficients (loadings). This process emphasises the degree of relationship between the various factors and the original variables. Rotation of the factors is achieved by many methods, broadly split into two classes. Oblique techniques relax the factor axes, leading to correlated factors. More commonly used, however, are orthogonal techniques giving uncorrelated factors. The standard method is the varimax method, an orthogonal technique, which rotates the factor axes until the variance of the squared factor loadings is maximised. This occurs when each variable has high loadings on one or a few factors as opposed to moderate loadings on several. Thus a much clearer picture of which features each factor is highlighting is obtained.

2.3.4.1 Factor Analysis in the QSAR study

Factor analysis was performed on the physico-chemical descriptor data associated with fathead minnow and Microtox toxicities using the BMDP statistical software program 4M. The standard varimax rotation was applied, significant factors being obtained with an eigenvalue greater than one. The scores for each factor for the chemicals were entered into best subsets regression analysis, in the MINITAB ver 7.1 statistical software, against the relevant toxicity.

2.3.5 Canonical Correlation Analysis

When considering a data set with more than one biological end-point (as is common in many pharmaceutical, pesticide and toxicity studies), it may be of value to investigate the relationship for all the biological data available for a set of compounds with respect to their corresponding descriptors. Multiple regression analysis can cope only with one dependent variable, so canonical correlation analysis has been utilised to relate the effect of the biological activities to their physico-chemical descriptors. Canonical correlation considers both sets of data and calculates linear, orthogonal, combinations of each, with a maximum correlation between them. The linear combinations of the data are known as 'canonical variates', and the correlation between each variate as the 'canonical correlation' (Harris, 1975). This process is similar to principal component analysis except here a correlation is maximised instead of a variance. The process of canonical correlation analysis is possibly best exemplified through the data available in this study.

2.3.5.1 Canonical Correlation Analysis in the QSAR Study

Canonical correlation analysis was performed using the BMDP statistical software program 6M. The first data set was the biological activities, comprising the fathead minnow and Microtox toxicities. The second data set was the physico-chemical descriptors of the chemials. In order to reduce the complexity of this large data matrix and thus increase the ease of understanding of the results, the fathead minnow 'decorrelated data' were utilised, namely CS3, PV1, ClogP, P(S-V)0, C(S-V)3, HOMO, MHOMO, LUMO, MLUMO and Mdipole.

2.3.6 Validation of the QSAR Models

The quantitative structure-activity relationships formed in any study should be tested in order to assess their robustness, and how useful they will be in the elucidation of the toxicities of chemicals for which no data are available. The approach used in this project is that of the testing set (see section 1.4.4) whereby the toxicities of chemicals for which data exist (but were not included in the original models) are calculated from the QSARs. The models tested were those formulated from regression analysis, principal component analysis and canonical correlation analysis for the toxicity of chemicals to the fathead minnow and the Microtox test.

2.3.6.1 Biological Data

Toxicity data for the fathead minnow 96 hour LC_{50}, and Microtox 5 minute EC_{50} were taken from the literature (Brooke et al, 1984; Geiger et al, 1985; 1986; 1988; Kaiser and Ribo, 1988). The chemicals chosen are listed in Tables 3.15 and 3.16. To test the validity of the regression analysis and the principal component analysis, one chemical was taken for each species from each chemical class analysed. For the canonical correlation analysis, chemicals were searched for that had toxicity data for both species. No such data were, however, found for the ketones, esters, and the nitriles, so four extra compounds for which both data were available, were analysed in addition.

2.3.6.2 Calculation of Physico-Chemical Parameters

For each compound the 49 physico-chemical parameters listed in Table 2.5 were calculated. This was performed according to the methods described in section 2.2.

2.3.6.3 Models Tested

Regression Analysis

For both species the toxicity was calculated initially from the best equation for all classes of chemical, and for the best equation (that complies with the Topliss and Costello (1972) rule) for the chemical class to which it belonged. The individual equations used are listed in Tables 3.15 and 3.16.

An estimate of the toxicity of a chemical was calculated simply by placing the

appropriate parameter(s) for the chemical into the relevant equation and calculating the toxicity.

Principal Component Analysis

The toxicities were calculated from the best equation found from the principal components. For each chemical the value for the separate principal components was obtained by finding the standarised value of each of the 49 parameters. Standardisation of the data was performed in the original analysis, and must be carried out in the validation process. Standardisation is achieved by:

$$\text{Standardised score} = \frac{\text{Score} - \text{Mean}}{\text{Standard Deviation}}$$

where the mean and standard deviation are the statistics for the original variable in the principal component analysis

The standardisation procedure eliminates the differing scale of variables, giving all variables a mean of zero, and a standard deviation of one.

The standardised scores were multiplied by the loadings of the principal component for each variable and summed to give the result for each principal component. The values for the principal components were then put into the regression equation.

N.B. It was not possible to utilise the results of the regression analysis on the factors, because the scores of the factors used in regression analysis had been standardised, and so were incompatible with unstandardised scores calculated for the testing set. (With principal component analysis unstandardised scores were used for the regression analysis.)

Canonical Correlation Analysis

Canonical correlation analysis (CCA) was used to extrapolate toxicity information from the relationship between the canonical variables for the best correlated variate, found in this project to be to be the first variate (eqn 4.7). Unlike the other predictive methods CCA is based on the prediction of two or more parameters, so one of these must be known for the equation to be solved. Initially the right hand side of the equation is solved, in a similar manner to

PCA. The variables were standardised, multiplied by the standardised coefficients for the canonical variable, and summed. The unknown toxicity was then found by placing the standardised fathead minnow, or Microtox, toxicity value in the equation to elucidate the other. (Unfortunately only four of the chemical classes were represented by the available data). The relevant toxicity was then standardised, multiplied by its loading and put into the equation so that it could be solved in the conventional manner.

2.4 Development of Inter-Species Relationships of Toxicity Between Four Aquatic Species

Extrapolation of toxicity data between species is a potentially useful method of assessing a chemical's hazard, as well as providing more toxicological information about the likely effects of a pollutant in the environment (see section 1.7). This part of the project investigates the inter-species relationships of toxicity between four aquatic species.

2.4.1. Data Compilation for Interspecies Relationships

A data base was compiled of the toxicities of 218 organic chemicals to the four commonly utilised aquatic toxicological endpoints described below. With the exception of 46 toxicities experimentally determined in the Microtox assay, all data have been extracted from the literature. N.B. This data base comprised the data used in the QSAR study (i.e that found in section 2.1) and has been expanded to include any other relevant and comparable toxicity data.

The relationships between the toxicity data are examined, with more detailed analysis of possible structural features or modes of toxic action that may be promoting greater toxic action to one species than to another.

2.4.1.1. Fathead Minnow Toxicity Data

These data were generated by the U.S. Environmental Protection Agency Research Laboratory, Duluth, in collaboration with the Center for Lake Superior Environmental Studies, University of Wisconsin-Superior, and are collated, with full methods, (Brooke et al 1984; Geiger et al, 1985; 1986; 1988). The toxicity data given are the 96 hour 50% lethality concentration for 30 to 35 day old fathead minnows (*Pimephales promelas*), determined by using flowthrough diluter systems at 25 $\pm 1^{o}C$. See section 2.1.2 for a more detailed description of the method.

2.4.1.2. Tetrahymena pyriformis Toxicity Data

These data have been obtained from the series of papers by Schultz and his coworkers (Schultz 1983; Schultz et al 1986; Moulton 1988; Cajina-Quezada 1988; Baker et al 1988; Dawson et al 1988; Schultz et al 1989a; Schultz et al 1989b).

The data are the 48 to 60 hour 50% inhibitory growth concentrations for the ciliate Tetrahymena pyriformis. These are determined using static axenic cultures of the ciliate maintained at 27 $\pm 1^{o}C$ in 250ml Erlenmeyer flasks, which contain 50 ml of medium (comprising distilled water, with 20 g/l proteose phosphate, 5 g/l glucose, 1 g/l yeast extract, 1 ml/l of a 3% w/v solution of Fe EDTA, and saturated NaOH to pH 7.35) innoculated with 0.25 ml of two day old log phase culture (approximately 36,000 cells per ml). The ciliates were grown in the culture, their population density measured spectrophotometrically as the optical density at 540 nm after 48 hours of incubation. (Concentration of T. pyriformis is directly proportional to the absorbance at 540 nm.) Each toxicant was assayed at least in triplicate.

2.4.1.3. Toxicity Data for the Microtox Test

The majority of the data, 117 of 163, are taken from a published compilation (Kaiser and Ribo, 1988). The data used are in each case the concentration that produces a 50% reduction in light output, from the marine bacterium Photobacterium phosphoreum after 5 minutes at 15 $\pm 0.1^{o}C$.

The remaining data were determined experimentally according to the procedure laid down by the manufacturers (Beckman Instruments Inc, 1982); again the concentration causing 50% reduction in light concentration after 5 minutes at 15 $\pm 0.1^{o}C$ was determined. See section 2.1.1. for full methods.

2.4.1.4. Toxicity Data for Daphnia magna

Many different toxicity data are available for Daphnia magna (e.g. effective, lethal, inhibition concentrations, no effect concentration, 24 and 48 hour, 21 day etc.). In order to obtain some consistency in this study only the concentration causing 50% lethality after 48 hours was used. However, even here slightly different methods have been used by different workers, (Adema 1978; Bobra et al 1983; Canton and Wegman 1983; Dill et al 1982; Eastmond et al 1984; Gersich et al 1986; Hermens et al 1984; LeBlanc 1980; LeBlanc 1984; Richter et al 1983; Sloof et al 1983; Thurston et al 1985). When more than one toxicity value

was available for each chemical the arithmetic mean was taken.

A general test protocol is described by LeBlanc (1980) thus. The desired concentrations of the chemical were added to 500 ml of distilled water in 2 litre jars. The 500 ml volume of test solution was then divided into three 150 ml aliquots in 250 ml beakers to provide triplicate exposures (the remaining 50 ml being used to assess dissolved oxygen concentration and pH). Five Daphnids were randomly placed in each 150 ml test solution. The solutions were maintained at 22 $\pm 1^{o}C$. Mortality data were collected after 48 hours for the range of concentrations, enabling a LC_{50} to be calculated.

2.4.2.1. Initial Statistical Manipulation of the Data

Once compiled, all of the data were converted into negative the logarithm of the millimolar concentration causing the described effects.

Initially regression analysis was performed on each 'pairing' of values, using the least-squares regression procedure of the MINITAB statistical package (ver 7.1). The six 'pairings' of data studied were:

i) Fathead Minnow vs Microtox
ii) Fathead Minnow vs *D. magna*
iii) Fathead Minnow vs *T. pyriformis*
iv) *D. magna* vs Microtox
v) *D. magna* vs *T. pyriformis*
vi) Microtox vs *T. pyriformis*

The chemicals were then sorted according to the presence of any of the 16 structural features listed in Table 2.6. These were considered to be all the important functional groups in the data set and are very similar to those proposed by Wallace and Niemi (1988) in a comparable study. The inter-species correlations were then reanalysed for each structural feature in order to investigate whether any feature led to a better correlation. This does of course mean that chemicals will be found in more than one of the separate correlations if more than one structural feature is present.

Table 2.6 Structural Features Analysed in the Inter-Species Toxcity Data

The presence of the following structural features was identified and noted for each chemical:

aromatic group	ester linkage
halogen	alcohol (including phenol)
aldehyde	ketone
carboxylic acid	amine (including aniline)
nitrile	N-ring (such as pyridine etc.)
nitro group	sulphur
phosphorus	alkene bond
heterocyclic group	ether

2.4.3.1. Identification of Outliers in the Inter-Species Relationships Involving Fathead Minnow Toxicity

The prediction of fathead minnow toxicity from another species was further investigated, to identify outliers having a significant effect on the inter-species relationships. From an analysis of the outliers it might, for instance, be possible to suggest features of a molecule that induce a greater relative toxic effect in one species as opposed to another. The outliers from these interspecies relationships were determined using the criterion that the predicted toxicity of the fathead minnow was greater, or less than, 4 times the experimentally measured fathead minnow toxicity i.e. the result of the equation

$$\frac{\text{Predicted fathead minnow toxicity}}{\text{Observed fathead minnow toxicity}}$$

was obtained for scalar data (as opposed to log transformed), and if the result was outside the limits of 0.25 to 4, the chemical was considered to be an outlier. This was the level of error applied by Wallace and Neimi (1988) in a similar study, and is, of course, purely arbitrary. The choice of 4 times the predicted value should also prevent any compounds appearing as outliers simply because of experimental error.

2.4.3.2. Analysis of the Outliers in the Inter-Species Relationships

The outliers from the interspecies relationships were analysed according to which structural features (listed in Table 2.6) they incorporated, in an attempt to deduce if any structural features of a molecule increased their relative toxicity

to one species as compared to another. The results are presented as a ratio:

$$\text{number of chemicals with greater toxicity to species A} \;:\; \text{number of chemicals with similar relative toxicity to A and B} \;:\; \text{number of chemicals with greater toxicity to species B}$$

calculated for all the chemicals in each relationship, and for the structural features being considered. In order to quantify whether the structural features are causing a significant increase in relative toxicity in one species, a one-way chi-squared analysis was performed on the data. The ratio of the number of chemicals with each of the relative toxicities for all the chemicals was considered to be the 'expected' ratio, and this was compared with the 'observed' (or experimental) ratio obtained for each structural feature using the following equation (Finney, 1980)

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

where O = number observed
E = number expected

There are 2 degrees of freedom.

2.4.4.1. Identification of Outliers from the Hydrophobicity and Toxicity Relationship for Each Species

Modern understanding of environmental QSAR is that chemicals may react by several modes of action (see section 1.6.3). Chemicals acting solely by a narcosis mechanism are well modelled by their hydrophobicity alone, unlike chemicals acting by more specific modes of action. Thus, assuming that purely narcotic chemicals are included in each relationship, a plot of toxicity versus hydrophobicity gives an indication of which chemicals may be acting by a narcosis mechanism, and which by a more specific mechanism.

The toxicities of the chemicals to each of the aquatic species were plotted against ClogP (as a measure of their hydrophobicity). A line representing the 'baseline toxicity' was then fitted to the graph, by eye, and chemicals more than

one log unit of toxicity above this line were assumed to be acting by a toxicity mechanism other than narcosis. The line representing baseline toxicity was positioned under the data points with the minimum toxicity (see, for example, Fig 3.6) so that it accounted for all chemicals acting by narcosis, it was felt that fitting the line by eye, would give a more accurate representation than fitting the line using regression analysis. The value of one log unit of toxicity prevents any compound being classified as an outlier due to experimental error, or error in estimating the partition coefficient. The position of each chemical on the graph was noted i.e. whether it is an outlier to the ClogP relationship (thus likely to be acting by a specific toxicity mechanism) or not. The compounds found to be outliers from the ClogP relationship were further analysed in each of the fathead minnow inter-species relationships. This information was used to investigate if a compound acting by a more specific toxicity mechanism to one species would make that species relatively more susceptible than another.

2.4.4.2. Analysis of the Outliers from the Toxicity-Hydrophobicity Relationship

A chi-squared analysis was performed on those chemicals identified as outliers from the toxicity-hydrophobicity (ClogP) relationship. These data were transformed into ratios so that for each species in the fathead minnow relationships, a two-way contingency table was created showing the number of ClogP outliers that were found to be significantly more toxic to one species than to another e.g.

	no. more toxic to species A	no. non-outlier	no. more toxic to species B
No. outliers from ClogP for Species A			
No. outliers from ClogP for Species B			

The two-way chi-squared test was performed using the MINITAB statistical software (ver 7.1). This form of test not only indicates whether consideration of the outliers from the hydrophobicity relationship is worthwhile, but which of the

Table 2.7 Summary of the Modes of Action Considered in Classifying the Outliers (Adapted from Hermens, 1989)

1. Narcosis - unreactive ketones and esters; some chlorinated alkanes and aromatic molecules; simple alcohols; sulphones; weak acids and bases; aliphatic nitrogenous compounds.

2. Uncouplers of oxidative phosphorylation - phenols (especially chlorophenols); dinitrophenols (especially 2,4 di-substituted); anilines and phenols with 2 or more substituted nitro groups or 4 or more ring substituted halogens.

3. Compounds metabolised to reactive intermediates - anilines; nitro substituted aromatics (especially dinitrobenzenes).

4. Compounds that may react with a nucleophile on a macromolecule - alkyl halides (especially the alkenes); epoxides; aldehydes; allylic and propargylic alcohols.

5. Polar Narcosis - Simple mono- and di- substituted phenols.

---

results are most significant.

2.4.5. Qualification of Outliers According to their Modes of Action

Various workers have established different modes of toxic action (see section 1.6.3), and these have been well reviewed by Hermens (1989). The outliers from the inter-species relationships were categorised according to their possible specific mode of toxic action. This may give an indication of which modes of action may cause greater relative toxic effects to one species as opposed to another. The modes of action considered are summarised in Table 2.7.

2.4.6.1. Improvement of Inter-Species Relationships Involving the Fathead Minnow using Physico-Chemical Parameters

In an attempt further to improve the predictability of fathead minnow toxicity, physicochemical parameters were calculated and included in a stepwise regression. Data included were the indicator variables for the presence of 16 structural

features of the molecules listed in Table 2.6; zero to third order path and valence-corrected molecular connectivities (as described in section 2.2.2); and ClogP (as described in section 2.2.1).

Each interspecies relationship involving fathead minnow toxicities was then re-analysed using forward stepwise regression, with fathead minnow toxicity as the dependent variable, and the other toxicity and the physico-chemical data as the independent variables. In each case the toxicity used as the predictor was 'forced' into the equation as the first independent variable, and a normal stepwise regression analysis followed.

The stepwise regression analysis was performed using the MINITAB statistical software (ver 7.1), and the variables chosen confirmed using best subsets regression analysis again in the MINITAB package.

2.4.6. Identification of the Most 'Representative' Test Species

The most representative test species is sought that will allow accurate estimations of a chemicals' environmental hazard. In an attempt to find this test species, from the four aquatic species already described, the essential criterion is that the ideal species will predict the toxicity of another accurately, or at least without significantly underestimating the risk, i.e. it is better for a toxicity prediction to show the hazard to be greater than it actually is, rather than for it to be predicted as less.

In order to quantify the process the residuals from the regression equations were analysed. The residuals give an indication of how divergent the estimate of toxicity is from the measured value. The residual is defined as:

$$\text{Residual} = \text{Measured Toxicity} - \text{Predicted Toxicity}$$

Thus residuals with a positive value indicate deviation from the regression equation that predict toxicity less than the measured value. N.B. In this case the negative logarithm of molar toxicity is being considered. These positive residuals were summed and mean per number of data points was calculated for each species acting as the predictor. (The mean was taken because there are a different number of chemicals used in each relationship.)

3. RESULTS

3.1 Biological Activities

3.1.1 Experimentally Determined Microtox Data

The mean and the standard error of the 5 and 15 minute EC_{50} experimentally determined data for the Microtox bioassay are listed in Table 3.1. The data are recorded as the negative logarithm of the millimolar concentration causing a 50% reduction in light output from the bacteria, as recommended by Kaiser and Ribo (1988).

The full results are listed in Appendix 3.

3.1.2. Other Biological Data

The whole biological data set used in the QSAR analysis is listed in Table 3.2. The Microtox data not experimentally obtained were taken from a data compilation by Kaiser and Ribo (1988). All fathead minnow data were extracted from the literature (Brooke et al, 1984; Geiger et al, 1985; 1986; 1988). For the purposes of consistency all data considered were the negative logarithm of the millimolar concentration causing the required effect.

Table 3.1 Mean and Standard Error of the Results of the Microtox Bioassay

Chemical	5 min $\log(1/EC_{50})$ mean	S.E.	15 min $\log(1/EC_{50})$ mean	S.E.
3-pentanone	-1.32	0.033	-1.39	0.030
5-nonanone	0.75	0.024	0.61	0.023
3-methyl-2-butanone	0.11	0.020	0.03	0.023
3,3-dimethyl-2-butanone	1.52	0.061	1.47	0.066
2-ethoxyethyl acetate	-0.93	0.032	-1.00	0.012
methyl acetate	-2.21	0.033	-2.18	0.000
propyl acetate	-0.49	0.013	-0.58	0.020
butyl acetate	0.06	0.034	-0.09	0.022
hexyl acetate	1.19	0.023	1.09	0.029
ethyl hexanoate	0.53	0.022	0.39	0.043
diethyl adipate	0.82	0.032	0.76	0.034
dibutyl adipate	1.94	0.076	1.94	0.090
diethyl sebacate	2.62	0.033	2.65	0.032
dimethyl malonate	-1.85	0.032	-1.83	0.023
chloroacetonitrile	-0.99	0.010	-0.46	0.020
malononitrile	-0.77	0.008	-0.36	0.003
allyl cyanide	-1.60	0.029	-1.52	0.015
1,4-dicyanobutane	-1.40	0.081	-1.49	0.095
1,6-dicyanohexane	0.93	0.010	0.84	0.006
octylcyanide	2.08	0.029	2.20	0.036
acetone	-2.56	0.026	-2.54	0.025
toluene	0.61	0.082	0.54	0.080
2-methoxyethylamine	0.48	0.046	0.55	0.048
1,2-diaminopropane	0.47	0.022	0.53	0.020
butanal	-0.57	0.045	-0.41	0.026
propylamine	0.67	0.040	0.81	0.013
2-chloro-4-methylaniline	1.43	0.025	1.37	0.030
octylamine	0.65	0.009	0.68	0.012
hexanal	0.51	0.040	0.67	0.034
heptylamine	0.63	0.011	0.68	0.026
4-fluoroaniline	0.13	0.067	0.095	0.062
N,N-diethylaniline	1.36	0.032	1.29	0.035
2-fluorobenzaldehyde	0.81	0.026	0.80	0.023
2-chloro-6-fluorobenzaldehyde	0.75	0.021	0.87	0.032
5-bromosalicylaldehyde	1.35	0.020	1.47	0.029
vanillin	0.42	0.020	0.36	0.054
2,4-dichlorobenzaldehyde	1.52	0.034	1.53	0.035
4-chloro-3-nitrotoluene	1.53	0.043	1.51	0.044
1,2,4-trichlorobenzene	2.04	0.044	1.97	0.023
2-chloronitrobenzene	1.58	0.029	1.56	0.027
3-chloronitrobenzene	1.17	0.030	1.15	0.027
2-chloro-4-nitrotoluene	1.71	0.041	1.65	0.020
2-chloro-6-nitrotoluene	2.38	0.012	2.33	0.012
acrolein	2.24	0.051	2.66	0.057
biphenyl	1.70	0.067	1.68	0.074
1,3-dichloropropanol	-1.16	0.057	-1.11	0.017
3-chlorotoluene	1.67	0.095	1.62	0.092
4-chloronitrobenzene	0.84	0.075	0.79	0.070

Table 3.2 Chemicals and their Fathead Minnow and Microtox Toxicities used in the QSAR Analysis

Chemical	Microtox 5 min $\log(1/EC_{50})$	Fathead Minnow 96 hr $\log(1/LC_{50})$
Ketones		
3,3-dimethyl-2-butanone	1.52	0.06
3-methyl-2-butanone	0.11	-1.00
3-pentanone	-1.32	-1.25
5-nonanone	0.75	0.66
acetone	-2.56	-2.09
2-butanone	-1.85	-1.65
2-octanone	0.86	0.55
2-decanone	1.30	1.50
2,4-pentandione	-1.02	-0.02
5-methyl-2-hexanone	-0.93	-0.14
4-methyl-2-pentanone	0.10	-0.72
6-methyl-5-heptene-2-one	0.86	0.17
cyclohexanone	0.72	-0.81
Esters		
dimethyl malonate	-1.85	1.03
2-ethoxyethylacetate	-0.93	0.46
methyl acetate	-2.21	-0.64
propyl acetate	-0.49	0.23
butyl acetate	0.06	0.81
hexyl acetate	1.19	1.55
ethyl caproate	0.53	1.21
dibutyl adipate	1.94	1.85
diethyl sebacate	2.62	1.98
diethyl adiapte	0.82	1.05
ethyl acetate	-1.77	-0.42
dibutyl succinate	-	1.71
diethyl malonate	-	1.02
diethyl chloromalonate	-	2.31
diethyl benzyl malonate	-	1.66
dibutyl fumarate	-	2.51
Nitriles		
1,4-dicyanobutane	-1.40	-1.25
allyl cyanide	-1.60	-0.43
chloroacetonitrile	-0.99	1.75
1,6-dicyanohexane	0.93	-0.59
malononitrile	-0.77	2.07
octyl cyanide	2.08	1.43
acetonitrile	-2.77	-1.60
undecyl cyanide	-	2.62
Aldehydes		
2,4-dichlorobenzaldehyde	1.52	1.99
2-chloro-6-fluorobenzaldehyde	0.75	1.23
2-fluorobenzaldehyde	0.81	1.96
5-bromosalicylaldehyde	1.35	2.19

butanal	-0.57	0.69
hexanal	0.51	0.85
vanillin	0.42	0.23
acrolein	2.24	3.46
Amines		
2-methoxyethylamine	0.48	-0.84
1,2-diaminopropane	0.47	-1.13
propylamine	0.67	-0.72
2-chloro-4-methyl aniline	1.43	0.60
octylamine	0.65	1.40
heptylamine	0.63	0.72
4-fluoroaniline	0.13	0.82
N,N-diethylaniline	1.36	0.96
Substituted Benzenes		
4-chloro-3-nitrotoluene	1.53	-
1,2,4-trichlorobenzene	2.04	-
2-chloronitrobenzene	1.58	-
3-chloronitrobenzene	1.17	0.92
2-chloro-4-nitrotoluene	1.71	-
2-chloro-6-nitrotoluene	2.38	-
3-chlorotoluene	1.67	-
4-chloronitrobenzene	0.84	-
toluene	0.61	0.41
biphenyl	1.70	-
Alcohols		
1,3-dichloro-2-propanol	-1.16	-
2-methyl-1-propanol	-1.35	-1.28
1-chloro-2-propanol	-	-0.41
2,2,2-trichloroethanol	-1.08	-0.30
2,3-dibromopropanol	-0.17	0.49
cyclohexanol	-0.06	-0.85
2-methyl-2,4-pentanediol	-1.41	-1.96
2-phenoxyethanol	-	-0.40
2-chloroethanol	-2.22	0.17
3-chloro-1-propanol	-	-0.93
methanol	-3.59	-2.94
ethanol	-2.98	-2.49
propanol	-2.47	-1.88
2-propanol	-2.77	-2.24
butanol	-1.49	-1.37
hexanol	0.40	0.02
octanol	1.32	0.98
nonanol	-	1.40
decanol	-	1.82
undecanol	-	2.22
dodecanol	-	2.27

Both toxicity values are the negative logarithm of the millimolar concentration causing the described effect.

3.2. Quantitative Structure-Activity Relationship Analysis of the Biological Data

3.2.1 Cluster Analysis on Physico-Chemical Variables

3.2.1.1. Cluster Analysis on Physico-Chemical Variables of Compounds for which Microtox Toxicity Data are Available

The results of the analysis on the Microtox data are summarised in Table 3.3 and graphically expressed in Fig. 3.1. 14 clusters are produced at the 90% similarity level. Over half (27 of 49) variables are found in the first cluster. These are steric values, including the majority of the molecular connectivities, COSMIC steric values, and molecular weight. Also included is the whole molecule polarisability (since polarisability is proportional to molar refractivity, it can be considered as a steric term), and the electronic energy term from COSMIC. Descriptors thought to represent size but which are not in this cluster are: the closest approach; second and third order Kappa values (although Kier (1987) suggests these encode the shape or symmetry rather than size); differences in simple and valence corrected connectivities; and cluster connectivities (which describe the branching of molecules). Hydrophobicity, in the form of ClogP, has clustered by itself. The electronic terms calculated from molecular modelling have formed 5 clusters. The HOMOs and LUMOs from the different methods of calculation (CNDO and MNDO) are in different clusters, the LUMO, however, joining with the difference between HOMO. The two dipole moments clustered together. A set of random numbers was also included, and it is reassuring to note that these are clustered significantly away from the other variables.

3.2.1.2 Cluster Analysis on Physico-Chemical Variables of Compounds for which Fathead Minnow Toxicity Data are Available

The results, summarised in Table 3.4 and Fig. 3.2, show that at the 90% similarity level 10 significant clusters are formed. These are the same as for the Microtox data, except that the large steric group also includes the second and third order Kappa values, closest approach, and one of the cluster molecular connectivities.

Again the random numbers are not well clustered.

Table 3.3 Summary of the Cluster Analysis on Variables for the Microtox Data

Description of Cluster	Contents	Number in Cluster	Similarity	Variable chosen as Representative
1. STERIC	Mol Wt, electronic energy, path simple & valence, and sums of, molecular connectivity, polarisability, CMR, 0 and 1 order Kappa values, COSMIC steric values	27	92.7%	First order valence corrected molecular connectivity
2. STERIC	Closest approach	1		Closest approach
3. SHAPE/SYMMETRY	Second order Kappa values	2	99.0%	Second order Kappa
4. SHAPE/SYMMETRY	Third order Kappa values	2	99.2%	Third order Kappa
5. STERIC	Differences in simple and valence path molecular connectivity	4	92.4%	Difference between zero order simple and valence path molecular connectivity
6. HYDROPHOBIC	ClogP	1		ClogP
7. BRANCHING	Difference third order simple and valence cluster molecular connectivity	1		Difference in third order simple and valence cluster molecular connectivity
8. HOMO	CNDO calculated HOMO	1		CNDO calculated HOMO
9. HOMO	MNDO calculated HOMO	1		MNDO calculated HOMO
10. LUMO	CNDO calculated LUMO and difference between HOMO	2	96.8%	CNDO calculated LUMO

Table 3.3 cont'd

11. LUMO	MNDO calculated LUMO and difference between HOMO	2	90.4%	MNDO calculated LUMO
12. BRANCHING	Third order cluster simple and valence molecular connectivities and their sum	3	95.3%	Third order cluster molecular connectivity
13. DIPOLE MOMENT	CNDO & MNDO calculated Dipole Moment	2	95.3%	MNDO calculated Dipole Moment
14. RANDOM NUMBERS	Random numbers	1	19.8%	

Figure 3.1 Cluster analysis of the 49 variables associated with the Microtox toxicity data.

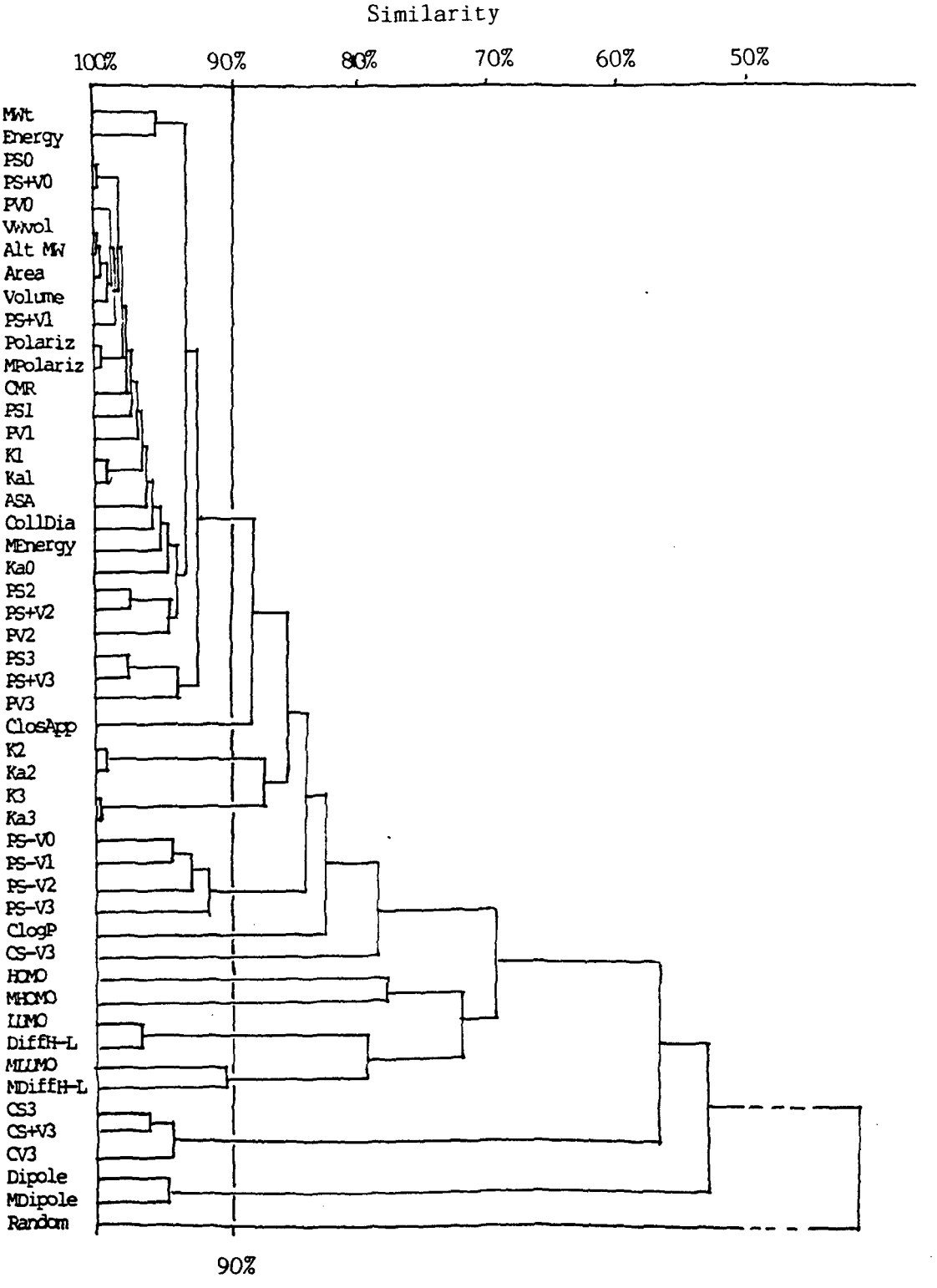


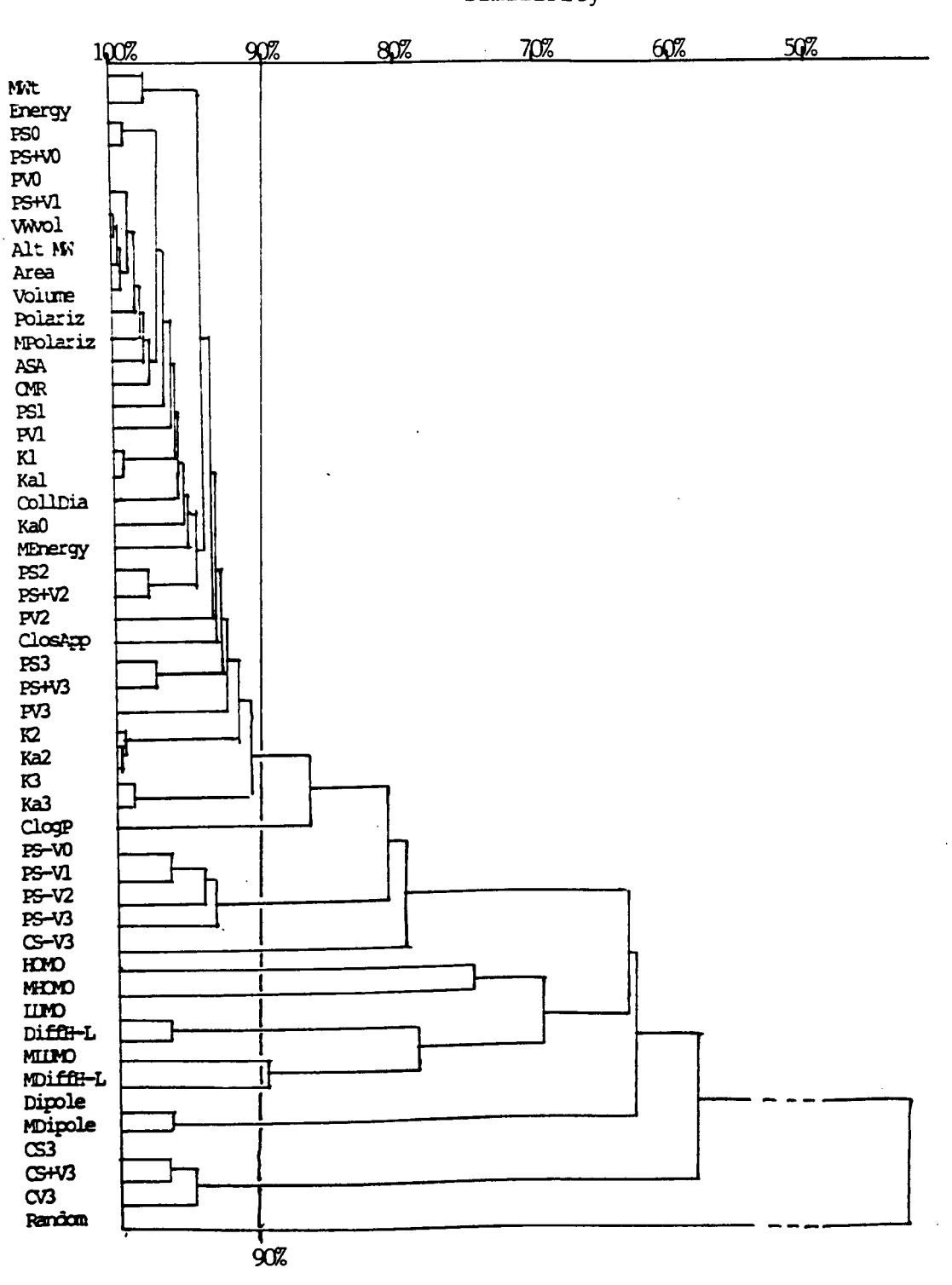
Table 3.4 Summary of the Cluster Analysis on Variables for the Fathead Minnow Data

Description of Cluster	Contents	Number in Cluster	Similarity	Variable chosen as Representative
1. STERIC	Mol Wt, electronic energy, path simple & valence, and sums of, molecular connectivity, polarisability, CMR, Kappa values, COSMIC steric values	32	90.6%	First order valence corrected molecular connectivity
2. HYDROPHOBIC	ClogP	1		ClogP
3. STERIC	Differences in simple and valence path molecular connectivity	4	93.7%	Difference between zero order simple and valence path molecular connectivity
4. BRANCHING	Difference third order simple and valence cluster molecular connectivity	1		Difference in third order simple and valence cluster molecular connectivity
5. HOMO	CNDO calculated HOMO	1		CNDO calculated HOMO
6. HOMO	MNDO calculated HOMO	1		MNDO calculated HOMO
7. LUMO	CNDO calculated LUMO and difference between HOMO	2	96.6%	CNDO calculated LUMO
8. LUMO	MNDO calculated LUMO and difference between HOMO	2	90.0%	MNDO calculated LUMO

Table 3.4 cont'd

9. DIPOLE MOMENT	CNDO & MNDO calculated Dipole Moment	2	95.3%	MNDO calculated Dipole Moment
10. BRANCHING	Third order cluster simple and valence molecular connectivities and their sum	3	96.4%	Third order cluster molecular connectivity
11. RANDOM NUMBERS	Random numbers	1	19.8%	

Figure 3.2 Cluster analysis of the 49 variables associated with fathead minnow toxicity data.



3.2.2. Stepwise Regression Analysis on Microtox Toxicities

The results of the correlation analysis of Microtox toxicity and decorrelated chemical descriptors are shown below

3.2.2.1. All Classes

$\log(1/EC_{50})_M = 0.805(0.081)\ ClogP - 1.13$ (3.1)

n=71 s=0.972 r^2(adj)=0.581 F=98.1

where $\log(1/EC_{50})_M$ is the inverse logarithm of the millimolar concentration causing a 50% reduction in light output in the Microtox test

ClogP is the logarithm of the calculated partition coefficient

$\log(1/EC_{50})_M = 0.631(0.090)\ ClogP + 0.495(0.14)\ HOMO + 5.69$ (3.2)

n=71 s=0.889 r^2(adj)=0.642 F=63.7

where HOMO is the CNDO calculated HOMO energy

$\log(1/EC_{50})_M = 0.525(0.099)\ ClogP + 0.546(0.14)\ HOMO + 0.104(0.047)\ K2 + 6.07$ (3.3)

n=71 s=0.874 r^2(adj)=0.661 F=46.6

where K2 is the second order Kappa value

These relationships were confirmed using best-subsets regression analysis.

3.2.2.2. Ketones

$\log(1/EC_{50})_M = 0.930(0.26)\ PV1 - 2.74$ (3.4)

n=13 s=0.924 r^2(adj)=0.489 F=12.5

where PV1 is the first order valence-corrected path molecular connectivity

$\log(1/EC_{50})_M = 2.52(0.59)\ PV1 - 0.805(0.28)\ K2 - 4.17$ (3.5)

n=13 s=0.719 r^2(adj)=0.690 F=14.4

3.2.2.3. Esters

$\log(1/EC_{50})_M = 1.11(0.108)\ ClogP - 1.96$ (3.6)

n=11 s=0.470 r^2(adj)=0.913 F=106

$\log(1/EC_{50})_M = 0.767(0.14)\ ClogP + 0.493(0.16)\ ClosApp - 4.00$ (3.7)

n=11 s=0.341 r^2(adj)=0.954 F=106

where ClosApp is the closest approach of another structure

$\log(1/EC_{50})_M$ = 0.708(0.098) ClogP + 1.01(0.20) ClosApp

- 1.97(0.64) HOMO - 32.8 (3.8)

n=11 s=0.238 r^2(adj)=0.978 F=148

3.2.2.4. Nitriles

$\log(1/EC_{50})_M$ = 1.060(0.24) PV1 - 2.96 (3.9)

n=7 s=0.820 r^2(adj)=0.749 F=18.9

N.B. A significant two parameter equation was not found.

3.2.2.5. Aldehydes

$\log(1/EC_{50})_M$ = - 0.599(0.32) MLUMO + 0.624 (3.10)

n=8 s=0.719 r^2(adj)=0.270 F=3.6

where MLUMO is the MNDO calculated LUMO energy

$\log(1/EC_{50})_M$ = - 1.68(0.37) MLUMO + 0.695(0.21) K3 - 1.89 (3.11)

n=8 s=0.436 r^2(adj)=0.731 F=10.5

where K3 is the third order Kappa value

3.2.2.5. Amines

$\log(1/EC_{50})_M$ = 0.152(0.090) ClogP + 0.508 (3.12)

n=8 s=0.398 r^2(adj)=0.208 F=2.8

$\log(1/EC_{50})_M$ = 0.169(0.091) ClogP - 0.0693(0.064) K3 + 0.771 (3.13)

n=8 s=0.392 r^2(adj)=0.229 F=2.0

3.2.2.6. Substituted Benzenes

$\log(1/EC_{50})_M$ = 0.768(0.32) PV1 - 0.928 (3.14)

n=10 s=0.427 r^2(adj)=0.348 F=5.8

3.2.2.7. Alcohols

$\log(1/EC_{50})_M$ = 1.30(0.12) PV1 - 4.29 (3.15)

n=14 s=0.443 r^2(adj)=0.898 F=116

$\log(1/EC_{50})_M$ = 0.956(0.093) PV1 + 0.485(0.092) ClogP - 3.81 (3.16)

n=14 s=0.247 r^2(adj)=0.968 F=200

3.2.2.8. Stepwise Regression Analysis on Fathead Minnow Toxicities

The results are shown below

All Classes

$$\log(1/LC_{50})_{FM} = 0.637(0.075)\ ClogP - 0.633 \quad (3.18)$$

n=75 s=0.999 r^2(adj)=0.486 F=71.0

where $\log(1/LC_{50})_{FM}$ is the inverse logarithm of the millimolar concentration causing 50% lethality in the fathead minnow

$$\log(1/LC_{50})_{FM} = 0.609(0.069)\ ClogP - 0.264(0.064)\ MLUMO - 0.224 \quad (3.19)$$

n=75 s=0.905 r^2(adj)=0.579 F=51.9

$$\log(1/LC_{50})_{FM} = 0.600(0.066)\ ClogP - 0.330(0.067)\ MLUMO - 0.700(0.28)\ CS3 + 0.146 \quad (3.20)$$

n=75 s=0.873 r^2(adj)=0.608 F=39.3

where CS3 is the third order simple cluster molecular connectivity

$$\log(1/LC_{50})_{FM} = 0.560(0.066)\ ClogP - 0.278(0.067)\ MLUMO - 0.861(0.27)\ CS3 + 0.221(0.83)\ MDipole - 0.350 \quad (3.21)$$

n=75 s=0.838 r^2(adj)=0.639 F=33.8

Equations 3.18 and 3.19 are confirmed by best-subsets regression analysis; however, when three and four variables are considered, the following relationships are revealed.

$$\log(1/LC_{50})_{FM} = 0.585(0.067)\ ClogP + 0.560(0.16)\ P(S-V)0 - 0.137(0.036)\ LUMO - 0.520 \quad (3.22)$$

n=75 s=0.870 r^2(adj)=0.611 F=39.7

where P(S-V)0 is the difference between simple and valence-corrected zero order path molecular connectivities
LUMO is the CNDO calculated LUMO energy

$$\log(1/LC_{50})_{FM} = 0.571(0.064)\ ClogP + 0.688(0.16)\ P(S-V)0 - 0.169(0.036)\ LUMO - 0.772(0.26)\ CS3 - 0.196 \quad (3.23)$$

n=75 s=0.826 r^2(adj)=0.649 F=35.2

3.2.2.9. Ketones

$$\log(1/LC_{50})_{FM} = 0.891(0.13)\ PV1 - 2.88 \quad (3.24)$$

n=13 s=0.464 r^2(adj)=0.788 F=45.5

$\log(1/LC_{50})_{FM}$ = 1.02(0.091) PV1 + 3.94(0.97) C(S-V)3 - 4.38 (3.25)

n=13 s=0.298 r^2(adj)=0.912 F=63.1

where C(S-V)3 is the difference between simple and valence-corrected third order cluster molecular connectivites

$\log(1/LC_{50})_{FM}$ = 1.05(0.061) PV1 + 2.99(0.70) C(S-V)3

+ 0.623(0.17) CS3 - 4.58 (3.26)

n=13 s=0.201 r^2(adj)=0.960 F=97.6

3.2.2.10. Esters

$\log(1/LC_{50})_{FM}$ = 0.555(0.098) ClogP - 0.0123 (3.27)

n=16 s=0.518 r^2(adj)=0.672 F=31.8

$\log(1/LC_{50})_{FM}$ = 0.447(0.092) ClogP + 0.584(0.22) P(S-V)0 - 0.714 (3.28)

n=16 s=0.435 r^2(adj)=0.768 F=25.8

3.2.2.11. Nitriles

$\log(1/LC_{50})_{FM}$ = 0.854(0.57) MDipole - 1.35 (3.29)

n=8 s=1.511 r^2(adj)=0.155 F=2.28

where MDipole is the MNDO calculated Dipole Moment

3.2.2.12. Aldehydes

$\log(1/LC_{50})_{FM}$ = 0.920(0.63) MDipole - 0.97 (3.30)

n=8 s=0.959 r^2(adj)=0.140 F=2.14

$\log(1/LC_{50})_{FM}$ = 1.01(0.60) MDipole - 0.109(0.084) LUMO - 1.11 (3.31)

n=8 s=0.909 r^2(adj)=0.227 F=2.03

3.2.2.13. Amines

$\log(1/LC_{50})_{FM}$ = 0.545(0.082) ClogP - 0.560 (3.32)

n=8 s=0.361 r^2(adj)=0.860 F=44.1

N.B. A significant two-parameter equation was not found.

3.2.2.14. Alcohols

$\log(1/LC_{50})_{FM}$ = 0.933(0.098) PV1 - 3.06 (3.33)

n=20 s=0.653 r^2(adj)=0.825 F=90.7

$\log(1/LC_{50})_{FM}$ = 0.915(0.082) PV1 - 1.41(0.48) C(S-V)3 - 2.94 (3.34)

n=20 s=0.547 r^2(adj)=0.878 F=69.1

3.2.3.1. Principal Component Analysis on Variables Associated with Microtox Data

Information concerning each principal component created by the principal component analysis is summarised in Table 3.5. There are 5 significant principal components with an eigenvalue greater than one. Between them, these 5 new uncorrelated variables have explained nearly 93% of the total variance of the original variables.

When the scores from the 5 principal components were put into best subsets regression analysis, the following relationships were formed:

$$\log(1/EC_{50})_M = 0.200(0.022)\ PC1 + 0.042 \quad (3.38)$$

$n=71 \quad s=1.03 \quad r^2(adj)=0.530 \quad F=79.8$

$$\log(1/EC_{50})_M = 0.200(0.021)\ PC1 - 0.252(0.071)\ PC4 + 0.042 \quad (3.39)$$

$n=71 \quad s=0.954 \quad r^2(adj)=0.597 \quad F=52.8$

$$\log(1/EC_{50})_M = 0.200(0.020)\ PC1 - 0.252(0.070)\ PC4 - 0.081(0.041)\ PC2 + 0.042 \quad (3.40)$$

$n=71 \quad s=0.934 \quad r^2(adj)=0.613 \quad F=38.0$

where PCn is the score for the n^{th} principal component

The loadings of the variables for each principal component are listed in Table 3.6.

3.2.3.2. Principal Component Analysis on Variables Associated with the Fathead Minnow Toxicity Data

Details of the principal components formed by the principal component analysis are summarised in Table 3.7. There are 5 significant principal components with an eigenvalue greater than 1, explaining almost 94% of the total variance of the original variables.

Best subsets regression analysis on the scores from the principal components reveals the following relationships:

$$\log(1/LC_{50})_{FM} = 0.167(0.021)\ PC1 + 0.342 \quad (3.41)$$

$n=75 \quad s=1.04 \quad r^2(adj)=0.444 \quad F=60.1$

$$\log(1/LC_{50})_{FM} = 0.167(0.020)\ PC1 - 0.310(0.084)\ PC5 + 0.342 \quad (3.42)$$

$n=75 \quad s=0.961 \quad r^2(adj)=0.525 \quad F=42.0$

$$\log(1/LC_{50})_{FM} = 0.167(0.018)\ PC1 - 0.310(0.078)\ PC5 - 0.248(0.067)\ PC4 + 0.342 \qquad (3.43)$$

$n=75 \qquad s=0.885 \qquad r^2(adj)=0.597 \qquad F=37.6$

The variable loadings are listed in Table 3.8.

Table 3.5 Summary of the Results of the Principal Component Analysis on the Variables for Microtox Data

Principal Component	Description	Eigenvalue	Proportion of Variance Explained	Cumulative Proportion Variance Explained	Significant Contributers and their Loading
1	Size and bulk	30.38	62.0%	62.0%	P(S+V)0 (0.180), PS0 (0.180), P(S+V)1 (0.178), PS1 (0.178), PS0 (0.178), CMR (0.178), Volume (0.178)
2	Electronic	7.31	14.9%	76.9%	MDiffH-L (-0.290, MLUMO (0.277), DiffH-L (-0.244), Ka3 (0.264), K3 (0.256), Ka2 (0.249), K2 (0.230)
3	Bulk and Branching	3.32	6.8%	83.7%	P(S-V)0 (0.400), P(S-V)1 (0.355), P(S-V)2 (0.302), CV3 (-0.375), C(S+V)3 (-0.289)
4	Branching	2.55	5.2%	88.9%	CS3 (0.356), C(S+V)3 (0.333), CV3 (0.257), C(S-V)3 (0.245) Dipole (0.329), MDipole (0.311), MHOMO (-0.280)
5	Electronic	1.88	3.8%	92.7%	LUMO (0.373), Dipole (-0.353), MDipole (-0.344), MHOMO (0.338), DiffH-L (-0.257)

Table 3.6 Eigenvectors (loadings) for the Significant Principal Components of the Variables Associated with the Microtox Data.

Variable	Principal Components				
	1	2	3	4	5
MWT	0.1674	-0.0484	-0.1020	-0.0297	-0.1560
PS0	0.1800	0.0038	0.0452	0.0359	0.0091
PS1	0.1784	0.0079	0.0757	-0.0454	0.0007
PS2	0.1719	-0.0953	0.0063	0.0629	0.0900
PS3	0.1601	-0.1280	0.0648	-0.1256	0.0075
CS3	0.0653	-0.2070	-0.1986	0.3568	0.2231
PV0	0.1781	0.0403	-0.0729	-0.0007	-0.0076
PV1	0.1727	0.0863	-0.0713	-0.0619	-0.0184
PV2	0.1617	-0.0028	-0.2257	0.0301	0.0613
PV3	0.1593	-0.0128	-0.1313	-0.1631	-0.0854
CV3	0.0228	-0.1573	-0.3757	0.2757	0.1561
KA0	0.1751	-0.0103	0.0604	-0.0047	-0.0426
K1	0.1743	0.0773	0.0417	0.0679	-0.0168
K2	0.1323	0.2301	0.0883	-0.0185	-0.1069
K3	0.0998	0.2565	-0.0192	0.1519	-0.0312
KA1	0.1709	0.1018	-0.0155	0.0828	-0.0315
KA2	0.1269	0.2490	0.0358	-0.0135	-0.1220
KA3	0.0925	0.2643	-0.0597	0.1539	-0.0502
CLOGP	0.1413	-0.0391	-0.0676	-0.2061	0.0094
CMR	0.1785	0.0132	-0.0283	-0.0715	0.0291
PS-V0	0.0862	-0.1159	0.4003	0.1341	0.0578
PS-V1	0.0965	-0.1667	0.3553	0.0145	0.0419
PS-V2	0.1156	-0.1760	0.3029	0.0799	0.0907
PS-V3	0.1121	-0.2042	0.2414	-0.0498	0.0981
CS-V3	0.0864	-0.1462	0.1963	0.2451	0.1777
PS+V0	0.1807	0.0210	-0.0101	0.0189	0.0013
PS+V1	0.1788	0.0440	0.0096	-0.0537	-0.0081
PS+V2	0.1734	-0.0597	-0.0908	0.0513	0.0810
PS+V3	0.1658	-0.0884	-0.0083	-0.1448	-0.0280
CS+V3	0.0482	-0.1925	-0.2897	0.3339	0.2012
ASA	0.1687	0.1178	-0.0118	-0.0027	0.0616
VWVOL	0.1764	0.0805	-0.0176	-0.0088	0.0309
ALTMV	0.1764	0.0805	-0.0177	-0.0088	0.0309
COLLDIA	0.1759	0.0547	-0.0408	-0.0228	0.0545
CLOSAPP	0.1475	0.1398	-0.0203	0.0741	0.0616
AREA	0.1744	0.0965	-0.0284	0.0115	0.0228
VOLUME	0.1781	0.0594	-0.0348	-0.0197	0.0173
DIPOLE	0.0674	-0.1203	0.0638	0.3294	-0.3531
ENERGY	-0.1629	0.0370	0.0502	-0.0188	0.1787
HOMO	0.1091	-0.1528	-0.0945	-0.1799	0.2284
LUMO	-0.0464	0.2363	0.1990	0.1377	0.3732
MDIPOLE	0.0724	-0.1236	0.1556	0.3118	-0.3449
MENERGY	-0.1728	0.0350	-0.0347	-0.0780	0.0786
MHOMO	0.0638	-0.1741	0.0203	-0.2805	0.3389
MLUMO	-0.0714	0.2777	0.0088	0.0504	0.2395
DIFH-L	0.0696	-0.2445	-0.1965	-0.1673	-0.2579
MDIFH-L	0.0832	-0.2902	0.0024	-0.1646	-0.0303
POLARIZ	0.1730	0.0888	0.0181	-0.0353	0.0827
MPOLARIZ	0.1710	0.0849	0.0442	-0.0413	0.1119

Table 3.7 Summary of the Results of the Principal Component Analysis on the Variables for the Fathead Minnow Data

Principal Component	Description	Eigenvalue	Proportion of Variance Explained	Cumulative Proportion Variance Explained	Significant Contributers and their Loadings
1	Size and bulk	31.5	64.3%	64.3%	P(S+V)0 (0.177), PS0 (0.177), PS1 (0.176), PV0 (0.176), P(S+V)1 (0.176), Volume (0.176)
2	Electronic	7.02	14.3%	78.6%	MDifH-L (-0.316), MLUMO (0.295), DiffH-L (-0.249), LUMO (0.233), CS3 (-0.243)
3	Bulk and Branching	3.25	6.7%	85.3%	P(S-V)0 (0.373), CV3 (-0.361), P(S-V)1 (0.330), P(S-V) (0.291), C(S+V)3 (-0.275), DiffH-L (-0.253), LUMO (0.241)
4	Branching	2.38	4.8%	90.1%	CS3 (0.393), C(S+V)3 (0.386), CV3 (0.341), MHOMO (-0.260), Dipole (0.244), DiffH-L (-0.231)
5	Electronic	1.75	3.6%	93.7%	MHOMO (0.433), LUMO (0.387), Dipole (-0.344), MDipole (-0.339), HOMO (0.283)

Table 3.8 Eigenvectors (loadings) for the Significant Principal Components of the Variables Associated with the Fathead Minnow Data.

Variable	Principal Components				
	1	2	3	4	5
MWT	0.1680	-0.0422	-0.0864	-0.0313	-0.1365
PS0	0.1770	-0.0156	0.0410	0.0283	-0.0021
PS1	0.1763	0.0000	0.0582	-0.0378	0.0040
PS2	0.1696	-0.0959	-0.0015	0.0675	0.0736
PS3	0.1618	-0.1057	0.0381	-0.1214	0.0416
CS3	0.0411	-0.2433	-0.1848	0.3937	0.1549
PV0	0.1762	0.0279	-0.0552	0.0153	-0.0122
PV1	0.1720	0.0803	-0.0621	-0.0328	-0.0040
PV2	0.1601	0.0284	-0.2163	0.0599	0.0578
PV3	0.1576	0.0486	-0.1552	-0.1391	-0.0106
CV3	0.0021	-0.1691	-0.3615	0.3411	0.0738
KA0	0.1739	0.0085	0.0293	-0.0496	0.0135
K1	0.1750	0.0327	0.0459	0.0521	-0.0360
K2	0.1513	0.1762	0.0717	-0.0309	-0.0895
K3	0.1289	0.1954	-0.0102	0.1251	-0.0726
KA1	0.1736	0.0550	0.0058	0.0683	-0.0478
KA2	0.1473	0.1970	0.0321	-0.0237	-0.0971
KA3	0.1226	0.2096	-0.0458	0.1287	-0.0865
CLOGP	0.1448	0.0719	-0.1128	-0.1603	0.0492
CMR	0.1758	0.0296	-0.0382	-0.0443	0.0393
PS-V0	0.0905	-0.1676	0.3732	0.0618	0.0361
PS-V1	0.0928	-0.2027	0.3306	-0.0301	0.0221
PS-V2	0.1046	-0.2208	0.2913	0.0469	0.0614
PS-V3	0.1049	-0.2284	0.2275	-0.0567	0.0809
CS-V3	0.0718	-0.1927	0.1994	0.2115	0.1729
PS+V0	0.1777	0.0048	-0.0041	0.0223	-0.0069
PS+V1	0.1766	0.0372	0.0032	-0.0360	0.0004
PS+V2	0.1714	-0.0456	-0.0943	0.0666	0.0693
PS+V3	0.1672	-0.0481	-0.0381	-0.1339	0.0224
CS+V3	0.0246	-0.2190	-0.2755	0.3860	0.1236
ASA	0.1726	0.0780	-0.0058	0.0172	0.0423
VWVOL	0.1751	0.0632	-0.0165	0.0058	0.0262
ALTMV	0.1751	0.0632	-0.0165	0.0058	0.0262
COLLDIA	0.1738	0.0455	-0.0446	-0.0096	0.0596
CLOSAPP	0.1629	0.0932	0.0001	0.0369	0.0658
AREA	0.1746	0.0701	-0.0178	0.0223	0.0138
VOLUME	0.1762	0.0480	-0.0270	0.0004	0.0124
DIPOLE	0.0850	-0.1332	0.1149	0.2442	-0.3441
ENERGY	-0.1633	0.0629	0.0323	0.0090	0.1574
HOMO	0.0967	-0.1681	-0.1567	-0.1884	0.2837
LUMO	-0.0203	0.2335	0.2415	0.2068	0.3873
MDIPOLE	0.0904	-0.1351	0.1854	0.2144	-0.3395
MENERGY	-0.1707	0.0553	-0.0450	-0.0470	0.0765
MHOMO	0.0392	-0.1981	-0.0202	-0.2601	0.4334
MLUMO	-0.0401	0.2955	0.0478	0.1416	0.2554
DIFH-L	0.0443	-0.2493	-0.2530	-0.2317	-0.2580
MDIFH-L	0.0488	-0.3169	-0.0457	-0.2287	0.0071
POLARIZ	0.1736	0.0678	0.0064	-0.0132	0.0720
MPOLARIZ	0.1723	0.0676	0.0241	-0.0156	0.0976

3.2.4.1. Factor Analysis on Variables Associated with the Microtox Toxicity Data

Information concerning the factors calculated by the factor analysis is summarised in Table 3.9; in addition, the loadings for the factors on the variables before and after varimax rotation are shown in Tables 3.10 and 3.11. Table 3.11 in fact shows the sorted rotated factors; in an attempt to make elucidation of the factors easier the loading factor matrix has been rearranged so that the columns appear in decreasing order of variance explained by each factor. The rows have been rearranged that so that for each successive factor, loadings greater than 0.50 appear first. Also loadings less than 0.25 have been replaced by zero. There are 5 significant factors, i.e. with an eigenvalue greater than one, and these explained 92.8% of the data space (and, of course, 100% of the factor space).

When the standardised scores from the 5 significant factors for each chemical were put into best-subsets regression, the following QSARs were obtained:

$$\log(1/EC_{50})_M = 0.936(0.14)\ FAC1 + 0.042 \qquad (3.44)$$

$n=71 \qquad s=1.184 \qquad r^2(adj)=0.379 \qquad F=43.7$

$$\log(1/EC_{50})_M = 0.936(0.12)\ FAC1 + 0.607(0.12)\ FAC3 + 0.042 \qquad (3.45)$$

$n=71 \qquad s=1.022 \qquad r^2(adj)=0.538 \qquad F=41.7$

$$\log(1/EC_{50})_M = 0.936(0.12)\ FAC1 + 0.607(0.12)\ FAC3 + 0.317(0.12)\ FAC2 + 0.042 \qquad (3.46)$$

$n=71 \qquad s=0.977 \qquad r^2(adj)=0.577 \qquad F=32.8$

$$\log(1/EC_{50})_M = 0.936(0.11)\ FAC1 + 0.607(0.11)\ FAC3 + 0.317(0.11)\ FAC2 - 0.294(0.11)\ FAC5 + 0.042 \qquad (3.47)$$

$n=71 \qquad s=0.937 \qquad r^2(adj)=0.611 \qquad F=28.5$

where: FACn is the score for the n^{th} rotated factor

3.2.4.2. Factor Analysis on Variables Associated with the Fathead Minnow Toxicity Data

The results of the factor analysis on the fathead minnow data are summarised in Table 3.12, and the unrotated and rotated loadings in Tables 3.13 and 3.14 respectively. There are 5 significant factors, explaining 93.7% of the variance

of the data space.

After best subsets regression analysis on the toxicity and the standardised factor scores, the following QSARs were revealed:

$$\log(1/LC_{50})_{FM} = 0.881(0.13)\ FAC1 + 0.342 \qquad (3.48)$$

n=75 s=1.088 r^2(adj)=0.391 F=48.5

$$\log(1/LC_{50})_{FM} = 0.881(0.11)\ FAC1 - 0.529(0.11)\ FAC3 + 0.342 \qquad (3.49)$$

n=75 s=0.917 r^2(adj)=0.530 F=42.8

$$\log(1/LC_{50})_{FM} = 0.881(0.11)\ FAC1 - 0.529(0.11)\ FAC3 - 0.285(0.11)\ FAC4 + 0.342 \qquad (3.50)$$

n=75 s=0.917 r^2(adj)=0.567 F=33.3

$$\log(1/LC_{50})_{FM} = 0.881(0.10)\ FAC1 - 0.529(0.10)\ FAC3 - 0.285(0.10)\ FAC4 + 0.221(0.10)\ FAC2 + 0.342 \qquad (3.51)$$

n=75 s=0.895 r^2(adj)=0.588 F=27.3

Table 3.9 Summary of the Results of the Factor Analysis on the Variables for Microtox Data

Rotated Factor	Description	Eigenvalue	Cumulative Proportion of Variance Explained in Data Sapce	in Factor Space	Significant Contributers and their Loading
1	Size and Bulk	26.2	53.5%	57.7%	Area (0.983), PV1 (0.976), ASA (0.974), VWvol (0.974), AltMW (0.974)
2	Size	7.06	67.9%	73.2%	P(S-V)0 (0.917), P(S-V)2 (0.909), P(S-V)1 (0.902), P(S-V)3 (0.821), C(S-V)3 (0.743)
3	Electron acception	6.07	80.3%	86.6%	LUMO (-0.933), DiffH-L (0.918), MLUMO (-0.789), MDiffH-L (0.757)
4	Branching	3.86	88.1%	95.0%	C(S+V)3 (0.963), CS3 (0.931), CV3 (0.916)
5	Electron donation	2.22	92.7%	100.0%	Dipole (0.689), MDipole (0.681), MHOMO (-0.674)

Table 3.10 Unrotated Factor Loadings for Variables Associated with the Microtox data.

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
MWT	0.923	-0.131	-0.186	-0.047	-0.214
PS0	0.992	0.010	0.082	0.057	0.012
PS1	0.983	0.021	0.138	-0.072	0.001
PS2	0.947	-0.258	0.012	0.101	0.123
PS3	0.883	-0.346	0.118	-0.201	0.010
CS3	0.360	-0.560	-0.362	0.570	0.306
PV0	0.982	0.109	-0.133	-0.001	-0.010
PV1	0.952	0.233	-0.130	-0.099	-0.025
PV2	0.891	-0.008	-0.412	0.048	0.084
PV3	0.878	-0.035	-0.239	-0.261	-0.117
CV3	0.125	-0.425	-0.685	0.440	0.214
KA0	0.965	-0.028	0.110	-0.008	-0.058
K1	0.961	0.209	0.076	0.109	-0.023
K2	0.729	0.622	0.161	-0.030	-0.147
K3	0.550	0.694	-0.035	0.243	-0.043
KA1	0.942	0.275	-0.028	0.132	-0.043
KA2	0.700	0.673	0.065	-0.022	-0.167
KA3	0.510	0.715	-0.109	0.246	-0.069
CLOGP	0.779	-0.106	-0.123	-0.329	0.013
CMR	0.984	0.036	-0.052	-0.114	0.040
PS-V0	0.475	-0.314	0.730	0.214	0.079
PS-V1	0.532	-0.451	0.648	0.023	0.058
PS-V2	0.637	-0.476	0.552	0.128	0.124
PS-V3	0.618	-0.552	0.440	-0.080	0.135
CS-V3	0.476	-0.395	0.358	0.392	0.244
PS+V0	0.996	0.057	-0.018	0.030	0.002
PS+V1	0.986	0.119	0.017	-0.086	-0.011
PS+V2	0.956	-0.161	-0.166	0.082	0.111
PS+V3	0.914	-0.239	-0.015	-0.231	-0.038
CS+V3	0.266	-0.521	-0.528	0.533	0.276
ASA	0.930	0.319	-0.021	-0.004	0.084
VWVOL	0.972	0.218	-0.032	-0.014	0.042
ALTMV	0.972	0.218	-0.032	-0.014	0.042
COLLDIA	0.969	0.148	-0.074	-0.036	0.075
CLOSAPP	0.813	0.378	-0.037	0.118	0.084
AREA	0.961	0.261	-0.052	0.018	0.031
VOLUME	0.982	0.161	-0.063	-0.031	0.024
DIPOLE	0.371	-0.325	0.116	0.526	-0.484
ENERGY	-0.898	0.100	0.091	-0.030	0.245
HOMO	0.601	-0.413	-0.172	-0.287	0.313
LUMO	-0.256	0.639	0.363	0.220	0.512
MDIPOLE	0.399	-0.334	0.284	0.498	-0.473
MENERGY	-0.952	0.095	-0.063	-0.125	0.108
MHOMO	0.351	-0.471	0.037	-0.448	0.465
MLUMO	-0.394	0.751	0.016	0.081	0.328
DIFH-L	0.383	-0.661	-0.358	-0.267	-0.354
MDIFH-L	0.458	-0.785	0.004	-0.263	-0.042
POLARIZ	0.954	0.240	0.033	-0.056	0.113
MPOLARIZ	0.942	0.230	0.081	-0.066	0.153
Eigenvalue	30.378	7.312	3.323	2.553	1.880

Table 3.11 Loadings for the Sorted Rotated Factors for the Variables Associated with the Microtox data.

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
AREA	0.983	0.000	0.000	0.000	0.000
PV1	0.976	0.000	0.000	0.000	0.000
ASA	0.974	0.000	0.000	0.000	0.000
VWVOL	0.974	0.000	0.000	0.000	0.000
ALTMV	0.974	0.000	0.000	0.000	0.000
KA1	0.964	0.000	0.000	0.000	0.000
VOLUME	0.964	0.000	0.000	0.000	0.000
POLARIZ	0.959	0.000	0.000	0.000	0.000
PV0	0.950	0.000	0.000	0.000	0.000
COLLDIA	0.949	0.000	0.000	0.000	0.000
K1	0.943	0.282	0.000	0.000	0.000
PS+V1	0.943	0.258	0.000	0.000	0.000
MPOLARIZ	0.939	0.276	0.000	0.000	0.000
PS+V0	0.927	0.288	0.000	0.000	0.000
CMR	0.917	0.252	0.253	0.000	0.000
KA2	0.899	0.000	0.000	-0.293	0.000
K2	0.894	0.000	0.000	-0.319	0.000
PS0	0.892	0.393	0.000	0.000	0.000
CLOSAPP	0.889	0.000	0.000	0.000	0.000
PS1	0.886	0.401	0.000	0.000	0.000
PV2	0.853	0.000	0.272	0.407	0.000
KA0	0.852	0.397	0.000	0.000	0.000
PV3	0.820	0.000	0.469	0.000	0.000
PS+V2	0.818	0.303	0.290	0.369	0.000
MENERGY	-0.814	-0.403	-0.269	0.000	0.000
MWT	0.809	0.000	0.484	0.000	0.000
ENERGY	-0.785	0.000	-0.413	0.000	0.000
K3	0.774	0.000	-0.400	0.000	0.000
KA3	0.755	0.000	-0.390	0.000	0.259
PS2	0.750	0.488	0.286	0.322	0.000
PS+V3	0.743	0.354	0.496	0.000	0.000
CLOGP	0.690	0.000	0.424	0.000	0.000
PS3	0.654	0.513	0.484	0.000	0.000
PS-V0	0.000	0.917	0.000	0.000	0.000
PS-V2	0.316	0.909	0.000	0.000	0.000
PS-V1	0.000	0.902	0.000	0.000	0.000
PS-V3	0.288	0.821	0.351	0.000	0.000
CS-V3	0.000	0.743	0.000	0.345	0.000
LUMO	0.000	0.000	-0.933	0.000	0.000
DIFH-L	0.000	0.000	0.918	0.000	0.000
MLUMO	0.000	-0.400	-0.789	0.000	0.000
MDIFH-L	0.000	0.482	0.757	0.000	0.000
CS+V3	0.000	0.000	0.000	0.963	0.000
CS3	0.000	0.285	0.000	0.931	0.000
CV3	0.000	0.000	0.000	0.916	0.000
DIPOLE	0.000	0.358	0.282	0.000	0.689
MDIPOLE	0.000	0.496	0.253	0.000	0.681
MHOMO	0.000	0.394	0.345	0.000	-0.674
HOMO	0.408	0.284	0.432	0.284	-0.476
Eigenvalue	26.226	7.056	6.073	3.865	2.223

Table 3.12 Summary of the Results of the Factor Analysis on the Variables for Fathead Minnow Data

Rotated Factor	Description	Eigenvalue	Cumulative Proportion of Variance Explained in Data Sapce	in Factor Space	Significant Contributers and their Loading
1	Size and Bulk	28.5	58.2%	61.2%	PV1 (0.991), Area (0.981), VWvol (0.979), AltMW (0.979), Volume (0.976)
2	Size	7.24	72.9%	77.9%	P(S-V)0 (0.928), P(S-V)2 (0.923), P(S-V)1 (0.905), P(S-V)3 (0.826), C(S-V)3 (0.750)
3	Electron acception	4.35	81.8%	87.4%	LUMO (0.962), DiffH-L (-0.937), MLUMO (0.802), MDiffH-L (-0.717)
4	Branching	3.50	89.0%	95.1%	C(S+V)3 (0.967), CS3 (0.931), CV3 (0.921)
5	Electron donation	2.28	93.6%	100.0%	MHOMO (0.809), HOMO (0.620)

Table 3.13 Unrotated Factor Loadings for the Variables Associated with the Fathead Minnow data.

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
MWT	0.943	0.112	-0.156	-0.048	-0.180
PS0	0.994	0.041	0.074	0.044	-0.003
PS1	0.990	0.000	0.105	-0.058	0.005
PS2	0.952	0.254	-0.003	0.104	0.097
PS3	0.908	0.280	0.069	-0.187	0.055
CS3	0.231	0.645	-0.333	0.608	0.205
PV0	0.989	-0.074	-0.100	0.024	-0.016
PV1	0.966	-0.213	-0.112	-0.051	-0.005
PV2	0.899	-0.075	-0.390	0.092	0.076
PV3	0.885	-0.129	-0.280	-0.215	-0.014
CV3	0.012	0.448	-0.652	0.527	0.098
KA0	0.976	-0.023	0.053	-0.077	0.018
K1	0.983	-0.087	0.083	0.080	-0.048
K2	0.850	-0.467	0.129	-0.048	-0.118
K3	0.724	-0.518	-0.018	0.193	-0.096
KA1	0.975	-0.146	0.010	0.105	-0.063
KA2	0.827	-0.522	0.058	-0.037	-0.128
KA3	0.688	-0.555	-0.083	0.199	-0.114
CLOGP	0.813	-0.190	-0.203	-0.247	0.065
CMR	0.987	-0.078	-0.069	-0.068	0.052
PS-V0	0.508	0.444	0.673	0.095	0.048
PS-V1	0.521	0.537	0.596	-0.047	0.029
PS-V2	0.587	0.585	0.525	0.072	0.081
PS-V3	0.589	0.605	0.410	-0.088	0.107
CS-V3	0.403	0.511	0.360	0.326	0.229
PS+V0	0.998	-0.013	-0.007	0.034	-0.009
PS+V1	0.992	-0.099	0.006	-0.056	0.001
PS+V2	0.962	0.121	-0.170	0.103	0.092
PS+V3	0.939	0.128	-0.069	-0.207	0.030
CS+V3	0.138	0.580	-0.497	0.596	0.163
ASA	0.969	-0.207	-0.011	0.026	0.056
VWVOL	0.983	-0.168	-0.030	0.009	0.035
ALTMV	0.983	-0.167	-0.030	0.009	0.035
COLLDIA	0.976	-0.121	-0.080	-0.015	0.079
CLOSAPP	0.914	-0.247	0.000	0.057	0.087
AREA	0.980	-0.186	-0.032	0.034	0.018
VOLUME	0.989	-0.127	-0.049	0.001	0.016
DIPOLE	0.477	0.353	0.207	0.377	-0.455
ENERGY	-0.917	-0.167	0.058	0.014	0.208
HOMO	0.543	0.445	-0.283	-0.291	0.375
LUMO	-0.114	-0.619	0.436	0.319	0.512
MDIPOLE	0.507	0.358	0.334	0.331	-0.449
MENERGY	-0.958	-0.146	-0.081	-0.073	0.101
MHOMO	0.220	0.525	-0.036	-0.401	0.573
MLUMO	-0.225	-0.783	0.086	0.219	0.338
DIFH-L	0.249	0.660	-0.456	-0.358	-0.341
MDIFH-L	0.274	0.840	-0.083	-0.353	0.009
POLARIZ	0.975	-0.180	0.011	-0.020	0.095
MPOLARIZ	0.967	-0.179	0.043	-0.024	0.129
Eigenvalue	31.512	7.020	3.253	2.382	1.747

Table 3.14 Loadings for the Sorted Rotated Factors for the Variables Associated with the Fathead Minnow data.

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
PV1	0.991	0.000	0.000	0.000	0.000
AREA	0.981	0.000	0.000	0.000	0.000
VWVOL	0.979	0.000	0.000	0.000	0.000
ALTMV	0.979	0.000	0.000	0.000	0.000
VOLUME	0.976	0.000	0.000	0.000	0.000
ASA	0.974	0.000	0.000	0.000	0.000
POLARIZ	0.969	0.000	0.000	0.000	0.000
PV0	0.966	0.000	0.000	0.000	0.000
COLLDIA	0.966	0.000	0.000	0.000	0.000
CMR	0.964	0.000	0.000	0.000	0.000
PS+V1	0.962	0.000	0.000	0.000	0.000
MPOLARIZ	0.958	0.000	0.000	0.000	0.000
KA1	0.956	0.000	0.000	0.000	0.000
PS+V0	0.941	0.304	0.000	0.000	0.000
K1	0.935	0.329	0.000	0.000	0.000
CLOSAPP	0.933	0.000	0.000	0.000	0.000
KA2	0.928	0.000	0.000	0.000	0.000
PV2	0.925	0.000	0.000	0.317	0.000
K2	0.922	0.000	0.000	-0.264	0.000
PV3	0.919	0.000	0.000	0.000	0.000
KA0	0.918	0.319	0.000	0.000	0.000
PS1	0.915	0.377	0.000	0.000	0.000
PS0	0.908	0.396	0.000	0.000	0.000
PS+V2	0.891	0.267	0.000	0.297	0.000
MWT	0.876	0.000	-0.356	0.000	0.000
CLOGP	0.860	0.000	0.000	0.000	0.000
PS+V3	0.859	0.284	-0.292	0.000	0.000
MENERGY	-0.841	-0.450	0.000	0.000	0.000
K3	0.836	0.000	0.000	0.000	-0.270
KA3	0.823	0.000	0.000	0.000	-0.289
ENERGY	-0.819	-0.327	0.347	0.000	0.000
PS2	0.815	0.464	0.000	0.265	0.000
PS3	0.762	0.468	-0.305	0.000	0.000
PS-V0	0.000	0.928	0.000	0.000	0.000
PS-V2	0.290	0.923	0.000	0.000	0.000
PS-V1	0.000	0.905	0.000	0.000	0.000
PS-V3	0.308	0.826	-0.257	0.000	0.000
CS-V3	0.000	0.750	0.000	0.330	0.000
MDIPOLE	0.306	0.643	0.000	0.000	-0.478
DIPOLE	0.298	0.544	-0.257	0.000	-0.490
LUMO	0.000	0.000	0.962	0.000	0.000
DIFH-L	0.000	0.000	-0.937	0.000	0.000
MLUMO	0.000	-0.406	0.802	0.000	0.000
MDIFH-L	0.000	0.452	-0.717	0.000	0.410
CS+V3	0.000	0.000	0.000	0.967	0.000
CS3	0.000	0.310	0.000	0.931	0.000
CV3	0.000	0.000	0.000	0.921	0.000
MHOMO	0.000	0.308	0.000	0.000	0.809
HOMO	0.426	0.000	-0.367	0.000	0.620
Eigenvalue	28.539	7.238	4.352	3.502	2.283

3.2.5.1. Canonical Correlation Analysis

62 corresponding fathead minnow and Microtox data were obtained and utilised in the analysis. Predictably the data are characterised by relatively high collinearity as expressed by the squared multiple correlations (r^2) of each variable in both the first and second data sets with all other in its corresponding data set, shown below. (Thus there is a correlation of 0.538 between the toxicities in the first data set; a correlation of 0.429 between CSl and all other members of the second data set and so on.)

Squared multiple correlation of each variable in the first and second sets with all other variables in their corresponding sets

First Set

Variable	r-squared
Microtox	0.538
Fathead Minnow	0.538

Second Set

CSl	0.429
PVl	0.848
ClogP	0.818
P(S-V)0	0.646
C(S-V)3	0.688
HOMO	0.853
LUMO	0.715
MDipole	0.413
MHOMO	0.792
MLUMO	0.823

3.2.5.2. Significance of the Canonical Variates

The maximum number of canonical variates (the linear combinations of the two canonical variables describing the first and second data sets) produced by the analysis is governed by the number of variables in the smaller of the two data sets. The first data set has only two variables, thus only two canonical variates can be calculated. Because some of the canonical correlations (the correlation between the canonical variables of the variate) may be too small to be statistically significant, Bartlett's Test is applied which indicates the number of significant relationships that exist. The results summarised below show that both the correlations of 0.837 for the first variate, and 0.507 for the second

variate, are significant at the 0.01% level.

Summary of statistics for canonical variates

No.	Eigenvalue	Canonical Correlation	Bartlett's Test for Significance Chi-square	Degrees of Freedom
1	0.701	0.837	104.3	20
2	0.507	0.712	38.5	9

The canonical correlation for each variate is simply, of course, the square root of the eigenvalue. Also Bartlett's test indicates that both variates are necessary to express the dependency between the two sets of variables. Both are significant at the 0.005% level, (the 0.005% chi-squared values being 40.0 for 20 degrees of freedom, and 23.6 for 9 degrees of freedom). These variables are, by their nature, uncorrelated.

3.2.5.3. Physical and Chemical Meaning of the Canonical Variates

The coefficients and standardised coefficients for the first and second set of canonical variables are shown below.

	Coefficients for canonical variables of both variates		Standardised coefficients for canonical variables of both variates	
First Set				
	CNVRF1	CNVRF2	CNVRF1	CNVRF2
Microtox	0.499	-0.869	0.732	-1.276
Fathead Minnow	0.246	1.068	0.330	1.433
Second Set				
	CNVRS1	CNVRS2	CNVRS1	CNVRS2
CS3	-0.277	-0.912	-0.114	-0.376
PV1	0.187	-0.375	0.244	-0.491
ClogP	0.305	0.349	0.410	0.469
P(S-V)0	0.547	0.497	0.313	0.284
C(S-V)3	-1.705	1.343	-0.373	0.293
HOMO	0.533	-0.601	0.504	-0.568
LUMO	-0.027	-0.209	-0.080	-0.611
MDipole	0.160	0.159	0.166	0.164
MHOMO	-0.067	-0.285	-0.070	-0.296
MLUMO	-0.001	-0.158	-0.002	-0.256

Where CNVRF1 is the canonical variable of the first set in the first variate
CNVRF2 is the canonical variable of the first set in the second variate
CNVRS1 is the canonical variable of the second set in the first variate
CNVRS2 is the canonical variable of the second set in the second variate

The standardised coefficients of the variates can be considered in a similar manner to the loadings of a principle component analysis, in order to elucidate the meaning of each variate. The first canonical variate is based on the sum of the weighted Microtox and weighted fathead minnow toxicities (basically in a ratio of 2:1). An increase in this variate is thus associated with an increase in both toxicities. The second canonical variate, however, describes the difference between the two toxicities (in roughly equal proportions) and an increase is associated with increase in fathead minnow toxicity, yet a decrease in Microtox toxicity.

There are relatively high canonical correlations between the first and second sets of variables for the variates. The first variate shows a good correlation of 0.837 (see Fig 3.3). CNVRF1 accounts for 94.9% of the variation of the Microtox toxicity and 75.2% of the fathead minnow toxicity. For the second canonical variate a correlation of 0.712 (see Fig 3.4) is obtained, with CNVRF2 explaining only 5.1% of the variation in the Microtox toxicity and 24.8% of that for the fathead minnow.

Figure 3.3 Graph of the scores of the canonical variables for the first (CNVRF1) and second (CNVRS1) data sets for the first canonical variate.

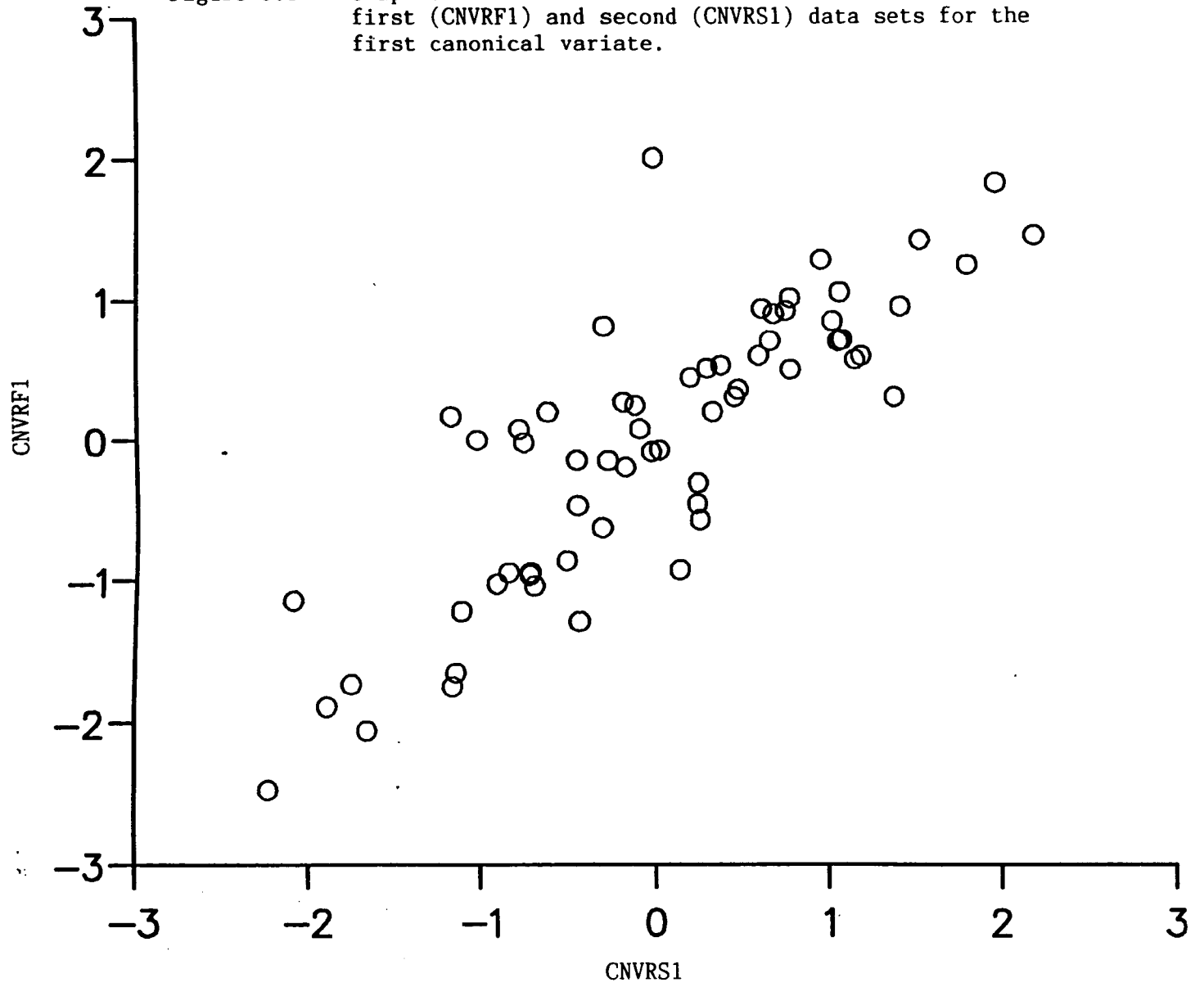
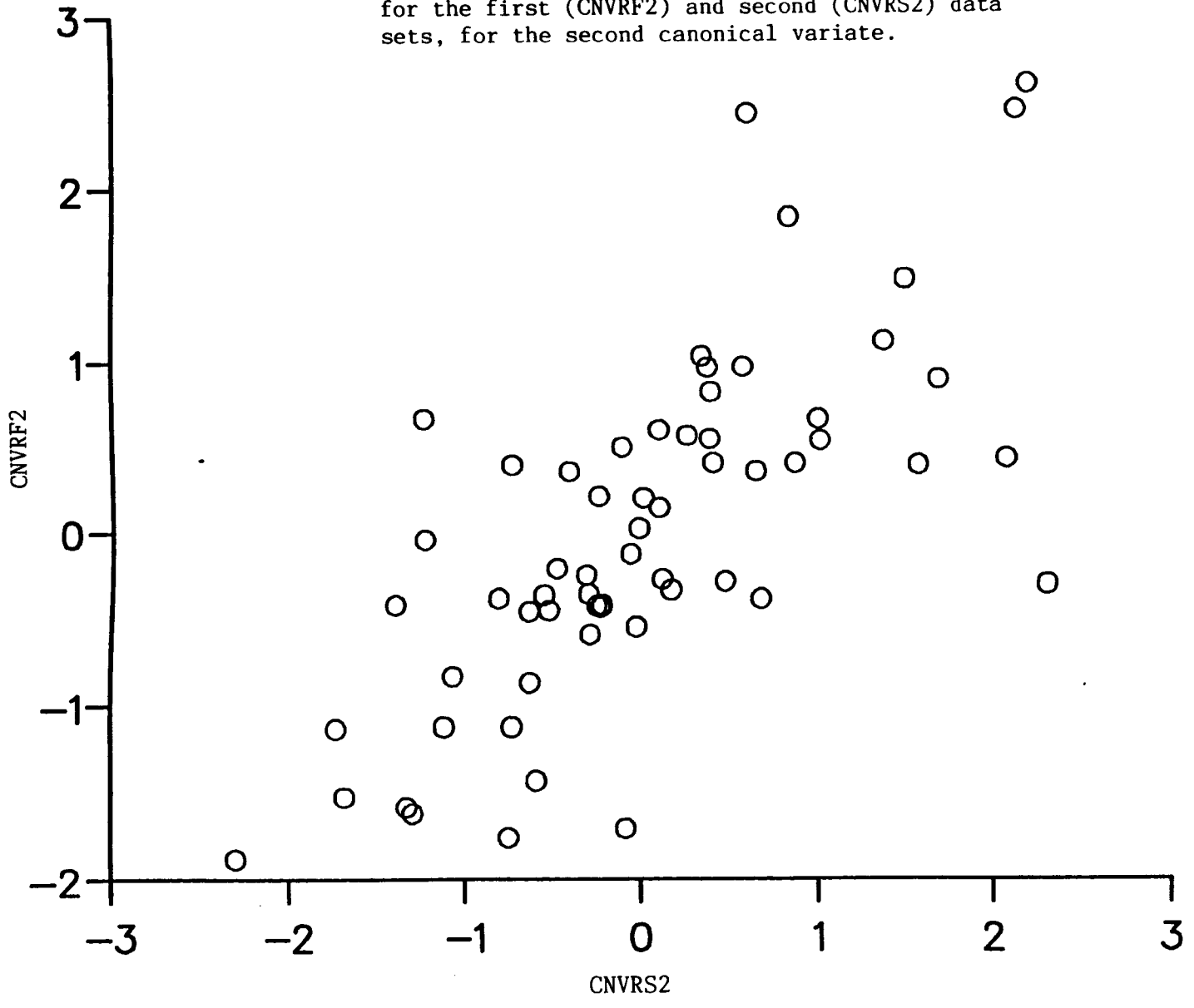


Figure 3.4 Graph of the scores of the canonical variables for the first (CNVRF2) and second (CNVRS2) data sets, for the second canonical variate.



3.2.6. Results of the Prediction of Toxicities from the QSAR Models

The chemicals in the 'testing set' for the QSAR analysis, along with their published toxicities, and those calculated by the various QSAR methods for the Microtox test are listed in Table 3.15, and those for the fathead minnow toxicity are listed in Table 3.16.

Table 3.15 Results of the Validation of Various QSAR Models for the Prediction of Microtox Toxicity

Chemical	Published Toxicity[1]	Calculated Toxicity				
		Regression Analysis All Classes[2]	Separate Class (eqn)[3]		Principal Component Analysis[4]	Canonical Correlation Analysis[5]
chloroacetone	0.53	-0.72	-1.29	(3.5)	0.85	-
ethylpropionate	-0.78	0.03	-0.71	(3.7)	1.39	-
4-chlorobenzonitrile	1.53	0.69	0.07	(3.9)	2.39	-
4-chlorobenzaldehyde	1.14	0.63	1.19	(3.10)	2.39	0.20
butylamine	-2.40	-0.78	0.65	(3.12)	1.22	-0.68
2,3-dichloronitro-benzene	2.17	0.90	1.74	(3.14)	2.58	-
heptanol	0.90	0.36	0.72	(3.16)	2.08	0.62
ethanal	-0.92	-	-		-	-1.38
aniline	0.36	-	-		-	-0.42
1,2-dichloroethane	-0.20	-	-		-	-1.33
1,3-dichlorobenzene	1.68	-	-		-	0.11
nitrobenzene	0.64	-	-		-	1.49

[1] taken from Kaiser and Ribo, 1988, and given as negative the logarithm of the millimolar concentration.
[2] calculated from equation 3.3.
[3] calculated from the equations listed below.
[4] calculated from equation 3.40.
[5] calculated from equation 4.7.

Table 3.16 Results of the Validation of Various QSAR Models for the Prediction of Fathead Minnow Toxicity

Chemical	Published Toxicity[1]	Calculated Toxicity				
		Regression Analysis All Classes[2]	Separate Class (eqn)[3]		Principal Component Analysis[4]	Canonical Correlation Analysis[5]
2-dodecanone	2.19	1.79	2.57	(3.25)	4.05	-
2,4-dihydroxymethyl-benzoate	0.56	1.72	2.69	(3.28)	2.72	-
acetonitrile	-1.60	-0.88	0.91	(3.29)	2.11	-
2-chloro-5-nitro-benzaldehyde	1.71	2.11	3.76	(3.31)	3.31	-
2-chloroaniline	1.35	0.73	0.49	(3.32)	2.31	-
heptanol	0.53	0.39	0.28	(3.34)	3.17	-0.04
4-chlorobenzaldehyde	1.80	-	-		-	-0.09
butylamine	-0.56	-	-		-	3.43
ethanal	0.16	-	-		-	-0.79
aniline	-0.16	-	-		-	1.33
1,2-dichloroethane	-0.14	-	-		-	-2.43
1,3-dichlorobenzene	1.26	-	-		-	-1.91
nitrobenzene	0.01	-	-		-	3.23

[1] taken from Brooke et al, 1984; Geiger et al, 1985; 1986; 1988; given as negative the logarithm of the millimolar concentration.
[2] calculated from equation 3.23.
[3] calculated from the equations listed below.
[4] calculated from equation 3.43.
[5] calculated from equation 4.7.

3.3. Results of the Analysis of the Inter-Species Relationships of Aquatic Toxicity

The inter-species relationships are described below, each relationship is then individually discussed in section 4.3.1., and an overview of the whole analysis given in section 4.3.2.

3.3.1.1. Inter-Species Relationship Between Fathead Minnow and Microtox

A total of 126 data pairings were obtained for the fathead minnow and Microtox toxicity data, and a reasonable relationship was found (see Fig 3.5):

$$\log(1/LC_{50})_{FM} = 0.704(0.046)\ \log(1/EC_{50})_{M} + 0.189 \qquad (3.52)$$

n=126 s=0.791 r^2adj=0.651 F=233.8

where: $\log(1/LC_{50})_{FM}$ = Fathead minnow 96h $\log(1/LC_{50})$
$\log(1/EC_{50})_{M}$ = Microtox 5 min $\log(1/EC_{50})$

When the data were broken down according to structural features present, much stronger relationships were observed in some cases:

alcohols:

$$\log(1/LC_{50})_{FM} = 0.834(0.059)\ \log(1/EC_{50})_{M} + 0.010 \qquad (3.53)$$

n=39 s=0.658 r^2adj=0.838 F=197.3

nitro compounds:

$$\log(1/LC_{50})_{FM} = 1.09(0.23)\ \log(1/EC_{50})_{M} - 0.260 \qquad (3.54)$$

n=8 s=0.391 r^2adj=0.755 F=22.5

esters:

$$\log(1/LC_{50})_{FM} = 0.702(0.12)\ \log(1/EC_{50})_{M} + 0.698 \qquad (3.55)$$

n=13 s=0.633 r^2adj=0.749 F=36.8

alkenes:

$$\log(1/LC_{50})_{FM} = 1.10(0.25)\ \log(1/EC_{50})_{M} + 0.412 \qquad (3.56)$$

n=7 s=0.902 r^2adj=0.746 F=18.7

The variation the slopes from (0.702 (esters) to 1.10 (alkenes)) and the intercepts (-0.260 (nitro compounds) to 0.698 (esters)) of these four equations indicate that different relationships occur when different chemical groups are considered. These differences also hint at the differing susceptibilities of the

species to toxicants.

Weaker, however, were the following relationships:

ethers:

$$\log(1/LC_{50})_{FM} = 1.10(0.26)\ \log(1/EC_{50})_{M} - 0.238 \qquad (3.57)$$

n=9 s=1.196 r^2adj=0.675 F=17.6

aldehydes:

$$\log(1/LC_{50})_{FM} = 0.887(0.19)\ \log(1/EC_{50})_{M} + 0.694 \qquad (3.58)$$

n=13 s=0.543 r^2adj=0.640 F=22.3

aliphatic compounds:

$$\log(1/LC_{50})_{FM} = 0.671(0.70)\ \log(1/EC_{50})_{M} + 0.143 \qquad (3.59)$$

n=68 s=0.868 r^2adj=0.576 F=91.9

halogenated compounds:

$$\log(1/LC_{50})_{FM} = 0.594(0.094)\ \log(1/EC_{50})_{M} + 0.655 \qquad (3.60)$$

n=38 s=0.731 r^2adj=0.511 F=39.7

aromatic compounds:

$$\log(1/LC_{50})_{FM} = 0.718(0.099)\ \log(1/EC_{50})_{M} + 0.214 \qquad (3.61)$$

n=58 s=0.699 r^2adj=0.475 F=52.6

ketones:

$$\log(1/LC_{50})_{FM} = 0.578(0.14)\ \log(1/EC_{50})_{M} - 0.158 \qquad (3.62)$$

n=19 s=0.720 r^2adj=0.458 F=16.2

nitrogen ring compounds:

$$\log(1/LC_{50})_{FM} = 0.329(0.16)\ \log(1/EC_{50})_{M} + 0.299 \qquad (3.63)$$

n=5 s=0.507 r^2adj=0.457 F=4.37

amines:

$$\log(1/LC_{50})_{FM} = 0.498(0.13)\ \log(1/EC_{50})_{M} + 0.191 \qquad (3.64)$$

n=25 s=0.822 r^2adj=0.380 F=15.7

nitriles:

$$\log(1/LC_{50})_{FM} = 0.380(0.30)\ \log(1/EC_{50})_{M} + 0.389 \qquad (3.65)$$

n=8 s=1.356 r^2adj=0.073 F=1.55

3.3.1.2. Analysis of the Outliers

The 'significant' (as defined in section 2.4.3.1) outliers from the inter-species relationship are listed in Table 3.17 and are divided into which species they are relatively more toxic against. Of the 126 chemicals, 78 (62%) are classified as not being significant outliers. Included in Table 3.17 are the possible modes of toxic action of the outliers (that are not assumed to be simple narcotics), as listed in Table 2.6.

3.3.1.3. Results of the Chi-Squared Analyis of the Outliers

The results of quantitative analysis into structural features causing greater relative toxicity to one species are summarised in Table 3.18. The overall ratio of all chemicals is 24 chemicals more toxic to the fathead minnow; 78 chemicals not relatively more toxic to either species; and 24 chemicals more toxic to Microtox (thus proportionally 19%:62%:19%). The chi-squared value shows that only when the aldehyde moiety is considered can the null hypothesis that there is no significant difference between the structural feature and the overall relationship be rejected (the chi-squared value for 95% probability of rejecting the null hypothesis, with 2 degrees of freedom, is 5.99). This suggests that the aldehyde group may be responsible for increased toxicity to the fathead minnow. However no other structural features are found to be significantly different from the expected ratio.

3.3.1.4. Results of the Analysis of Outliers from the Toxicity-Hydrophobicity Relationship

The outliers from the fathead minnow toxicity-ClogP (see Fig 3.6) and Microtox-ClogP (Fig 3.7) relationships (see section 2.4.4.1) are marked on Table 3.17. Of those chemicals found to be relatively more toxic to the fathead minnow, more (sixteen) are outliers to the fathead minnow-ClogP relationship, than to the Microtox-ClogP relationship (ten). Conversely, of those chemicals relatively more toxic in the Microtox test, more (twenty) are outliers to the Microtox-ClogP relationship, than are outliers to the fathead minnow-ClogP relationship (three). These results are quantified in Table 3.19 which shows a significant difference

between the two sets of ClogP outliers (the chi-squared value of 12.2 is greater than the 95% probability value to reject the null hypothesis of 5.99).

3.3.1.5. Improvement of the Inter-Species Relationship by Addition of Other Parameters

Following stepwise regression analysis with the physico-chemical data described in section 2.4.6.1. (ClogP, molecular connectivities, and indicator variables), Equation 3.52 was moderately improved:

$$\log(1/LC_{50})_{FM} = 0.490(0.069)\ \log(1/EC_{50})_{M} + 0.280(0.071)\ ClogP - 0.218 \qquad (3.66)$$

n=126 s=0.748 r^2adj=0.688 F=138

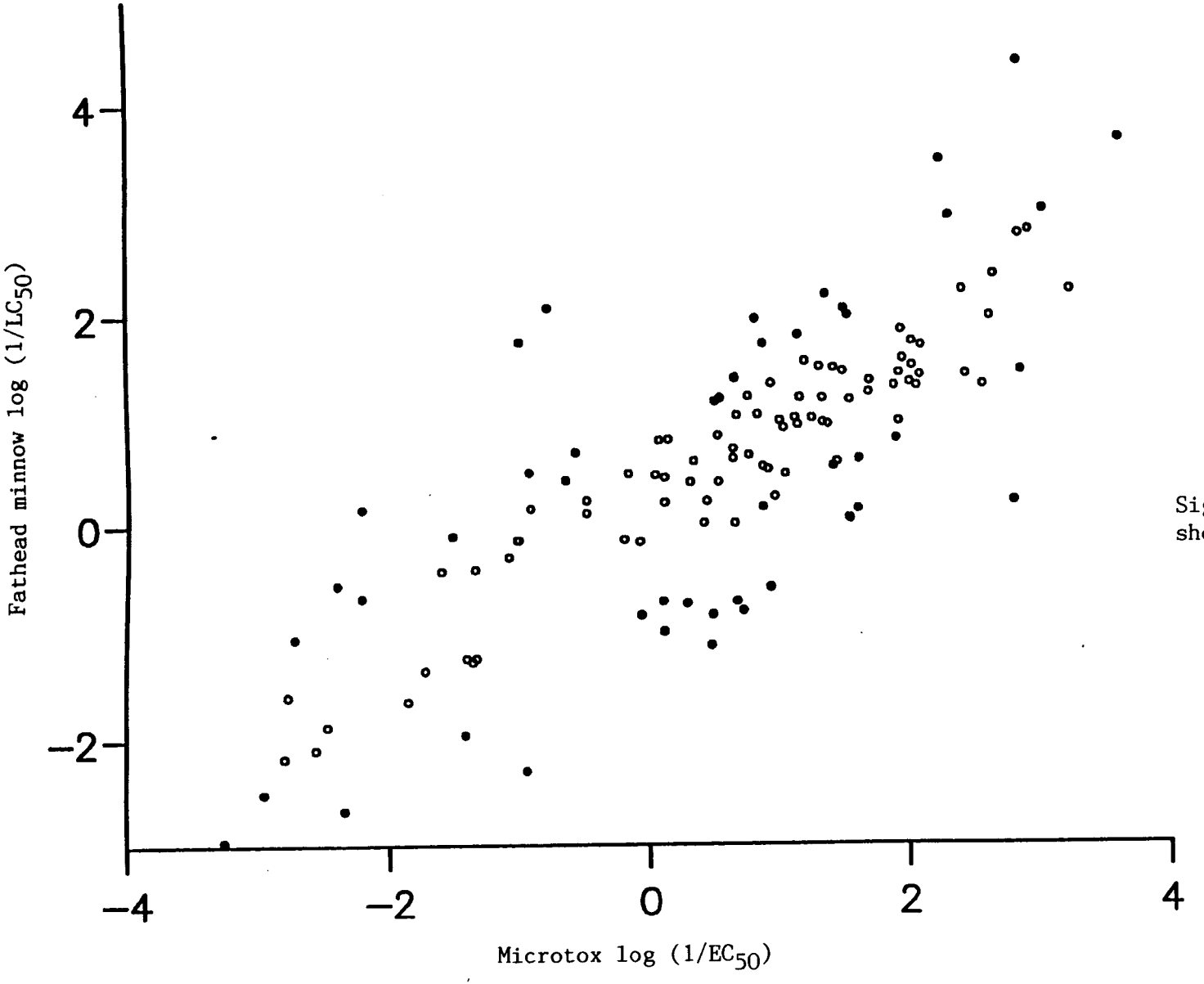
$$\log(1/LC_{50})_{FM} = 0.451(0.066)\ \log(1/EC_{50})_{M} + 0.319(0.067)\ ClogP + 0.867(0.21)\ I_{ald} - 0.358 \qquad (3.67)$$

n=126 s=0.703 r^2adj=0.724 F=110

where: I_{ald} is an indicator variable for the presence of an aldehyde group.

Equations 3.66 and 3.67 were the only significant improvements to the relationship that were obtained; the addition of more parameters did not improve the correlation appreciably. These results were confirmed by best-subsets regression analysis.

Figure 3.5 Graph of fathead minnow toxicity against Microtox toxicity.



Significant outliers are shown as solid circles.

Figure 3.6 Graph of fathead minnow toxicity against ClogP, showing outliers.

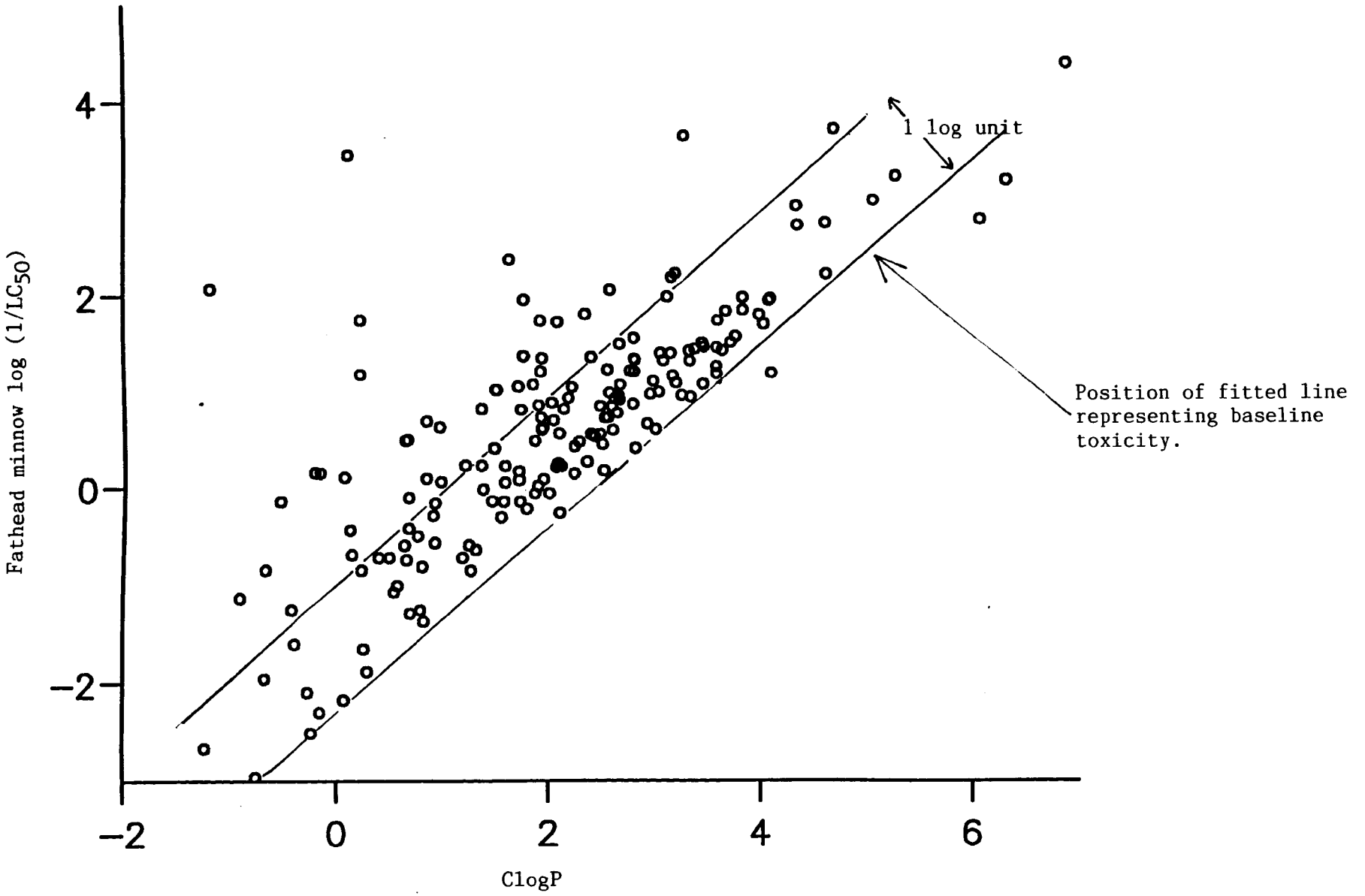


Figure 3.7 Graph of Microtox toxicity against ClogP, showing outliers.

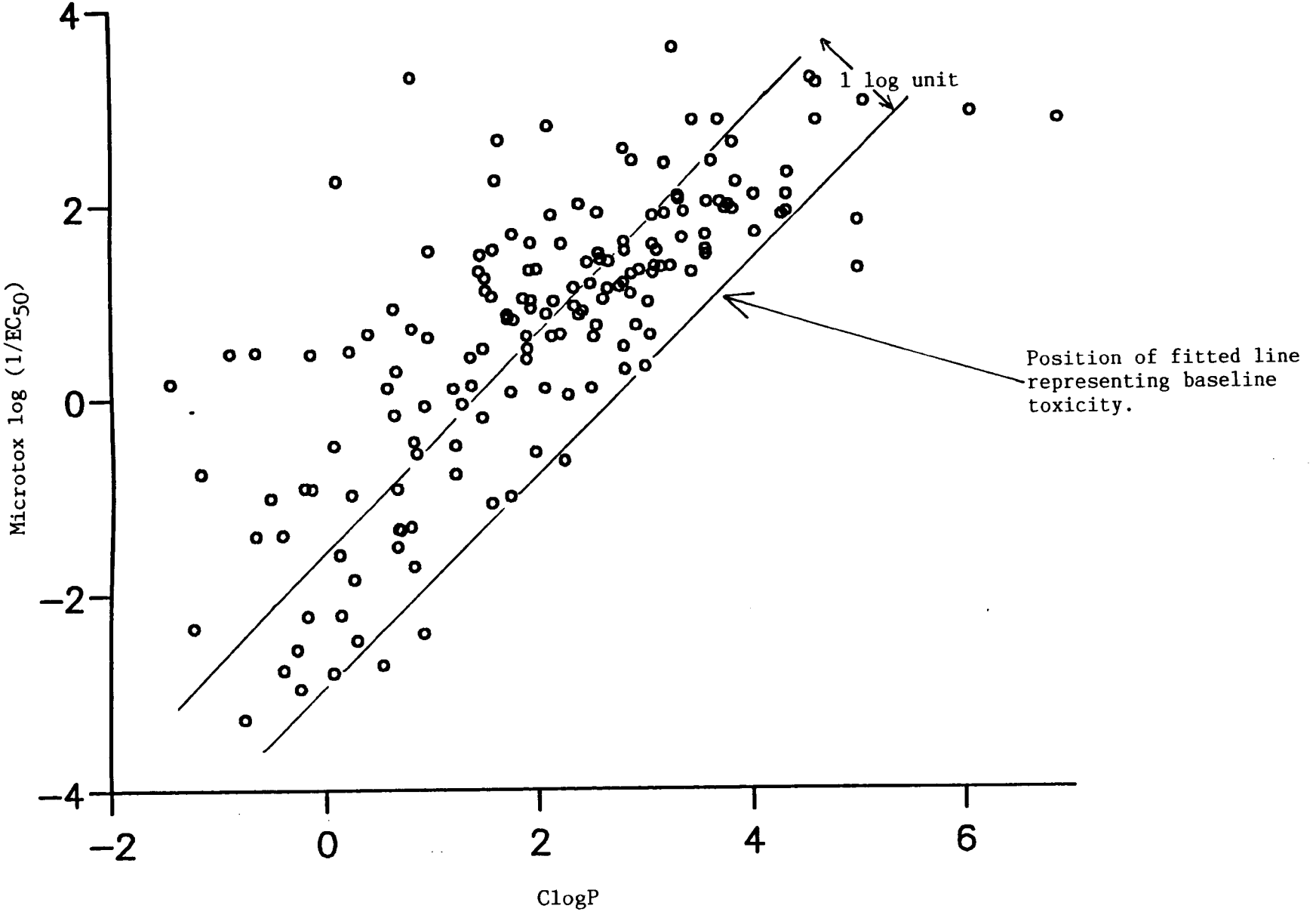


Table 3.17 Classification of Compounds the Fathead Minnow and Microtox Relationship

Compounds are ranked according to their relative toxicities; F and M after the chemical indicate that it is an outlier in the toxicity-ClogP relationship for fathead minnow and Microtox respectively; Modes indicate the possible modes of toxic action of the outliers (narcosis is not shown), where 2 are uncouplers of oxidative phosphorylation, 3 are compounds metabolised to reactive intermediates, 4 are compounds that may react with a nucleophilic group on a macromolecule, 5 are compounds acting by polar narcosis (see Table 2.7).

Chemical	Ratio Calc/Obs	ClogP Outlier		Mode
Chemicals More Toxic to Fathead Minnow				
malononitrile	0.0038	F	M	
chloroacetonitrile	0.0056	F	M	
permethrin	0.0063			
acrolein	0.02	F	M	2
2-chloroethanol	0.030	F		
2-fluorobenzaldehyde	0.063	F	M	4
2,3,5,6-tetrachloroaniline	0.08			2
5-bromosalicylaldehyde	0.089	F		4
2-ethoxyethyl acetate	0.11	F		
butanal	0.12	F	M	4
salicyladehyde	0.12	F	M	4
a,a'-dichloro-p-xylene	0.12	F	M	
butylamine	0.12			
4-chlorobenzaldehyde	0.15	F		4
4,6-dinitro-2-cresol	0.15	F	M	3
pyridine	0.17	F		
2,3,4,5-tetrachlorophenol	0.17			2
octylamine	0.18			
2,4-dichlorobenzaldehyde	0.19			4
amobarbital	0.20			
methyl acetate	0.21	F		
diethylamine	0.22			
4-dimethylamino-3-methyl-2-butanone	0.22	F	M	
pentachlorophenol	0.22			2
ethyl hexanoate	0.23			
ethanal	0.24	F	M	4
Non-Outliers				
butyl acetate	0.26	F		
dicofol	0.29			
4-fluoroaniline	0.29	F	M	
hexyl acetate	0.29			
2-chloroaniline	0.31	F	M	
2-chloro-6-fluorobenzaldehyde	0.31			
allyl cyanide	0.31	F		
2,3-dibromopropanol	0.38	F	M	
2-chlorophenol	0.40			
2-decanone	0.40			
propyl acetate	0.41	F		
2,4-pentanedione	0.41	F	M	
5-methyl-2-hexanone	0.42			
4-phenylazophenol	0.45	F	M	

ethyl acetate	0.46	F	
1,4-dinitrobenzene	0.48	F	M
hexanal	0.49	F	
1-naphthol	0.49		M
dibutyl adipate	0.51		
diethyl adipate	0.52	F	M
caffeine	0.54	F	M
2,2,2-trichloroethanol	0.54		
trichloroethylene	0.54		
2,4,6-trichlorophenol	0.59		
diuron	0.61		
butyl ether	0.64		
1,1,1-trichloroethane	0.65		
acetonitrile	0.69		
2,3,4-trichloroaniline	0.74		
tetrachloroethylene	0.78		
heptylamine	0.81		
2,4-dinitrophenol	0.81	F	M
2-nitrobenzaldehyde	0.90	F	M
2,4,6-tribromophenol	0.90		
3-chloronitrobenzene	0.96		
4-phenoxyphenol	0.97		
toluene	0.97		
4-amino-2-nitrophenol	1.01	F	M
malathion	1.02	F	M
2-allylphenol	1.08		
1,1,2-trichloroethane	1.11		
benzaldehyde	1.11	F	M
5-nonanone	1.13		
diethyl sebacate	1.13		
1,2-dichlorobenzene	1.19		
2-phenylphenol	1.25		
diazinon	1.25		
1,3-dichlorobenzene	1.29		
phenol	1.39		M
octanol	1.39		
4-butylaniline	1.54		
1,2-dichloroethane	1.54		
2,4-dichlorophenol	1.55		M
octyl cyanide	1.70		M
carbaryl	1.74	F	M
hexachloroethane	1.74		
2-octanone	1.75		
aniline	1.95		M
heptanol	1.97		
naphthalene	2.05		
propanol	2.14		
butanol	2.22		
2-propanol	2.43		
4-nitrophenol	2.70		M
hexanol	2.82		
1,4-dicyanobutane	2.85	F	M
pentachloroethane	2.95		M
acetone	3.02		
3-pentanone	3.25		
2-methyl-1-propanol	3.34		
2-butanone	3.44		
4-phenylpyridine	3.54		
N,N-dimethylaniline	3.89		

2-chloro-4-methylaniline	3.98		M	
Chemicals More Toxic to Microtox				
ethanol	4.08			
a,a,a-trifluoro-3-tolunitrile	4.17		M	
nitrobenzene	4.21		M	
6-methyl-5-heptene-2-one	4.23			
3,4-dichloroaniline	4.59		M	5
4-chloroaniline	4.96		M	
4-methylphenol	5.05		M	5
4-tert-butylphenol	5.38		M	
methanol	7.07			
4-methyl-2-pentanone	9.47		M	
cyclohexanol	9.85		M	
benzamide	13.27		M	
2-methyl-2,4-pentanediol	14.23		M	
bromacil	14.49		M	
3,3-dimethyl-2-butanone	15.77	F	M	
triethylene glycol	16.04		M	5
4-methoxyphenol	16.35		M	5 3-
methyl-2-butanone	18.53		M	
2-methoxyethylamine	23.47	F	M	
propylamine	23.85		M	
1,6-dicyanohexane	27.05			
cyclohexanone	31.82		M	
1,2-diaminopropane	45.10	F	M	
2-(2-ethoxyethoxy)ethanol	67.58		M	
4-ethylaniline	85.71		M	3

Table 3.18 Results of the Chi-Squared Analysis on the Outliers of the Fathead Minnow and Microtox Relationship by Structural Features of the Chemicals

The table indicates the number of chemicals with each of the following structural features and how many fall into each category. The expected ratio is calculated from the overall ratio.

	More toxic to fathead minnow	Non-outlier	More toxic to Microtox	Chi-squared value
Overall relationship				
	24	78	24	
Aromatic compounds				
Observed	12	37	9	
Expected	11	36	11	0.479
Esters				
Observed	3	8	3	
Expected	3	8	3	0.000
Halogenated compounds				
Observed	11	23	4	
Expected	7	24	7	3.60
Alcohols				
Observed	5	25	9	
Expected	7.5	24	7.5	1.70
Aldehydes				
Observed	8	5	0	
Expected	2.5	8	2.5	15.7
Ketones				
Observed	2	11	5	
Expected	3.5	11	3.5	1.28
Amines				
Observed	7	11	7	
Expected	5	15	5	2.67
Nitriles				
Observed	2	4	2	
Expected	1.5	5	1.5	0.54

Ethers				
Observed	1	4	3	
Expected	1.5	5	1.5	1.87

Table 3.19 Chi-Squared Analysis of the Outliers from the Toxicity-ClogP Relationship

		more toxic to Fathead Minnow	non-outliers	more toxic to Microtox		TOTAL
Fathead minnow-ClogP Outliers	O:	16	20	3		39
	E:	10.9	18.5	9.6		
Microtox-ClogP Outliers	O:	10	24	20		54
	E:	15.1	25.5	13.4		
Total		26	44	23		93
Chi-squared		2.383	0.130	4.578		
		1.721	0.094	3.307	=	12.2

where O: is the observed ratio
E: is the expected ratio calculated by the chi-squared analysis

3.3.2.1. Inter-Species Relationship between Fathead Minnow and Daphnia magna

A total of 46 data pairings were found for fathead minnow and D. magna, revealing the following relationship (see Fig 3.8):

$$\log(1/LC_{50})_{FM} = 0.805(0.069)\ \log(1/LC_{50})_{DM} + 0.061 \quad (3.68)$$

n=46 s=0.675 r^2adj=0.750 F=135.8

where: $\log(1/LC_{50})_{DM}$ is the log (1/(48 hour LC_{50})) to D. magna.

When the compounds were divided according to which structural features they contain, some very good correlations were obtained:

alcohols:

$$\log(1/LC_{50})_{FM} = 1.03(0.074)\ \log(1/LC_{50})_{DM} - 0.275 \quad (3.69)$$

n=17 s=0.415 r^2adj=0.923 F=193.2

alkenes:

$$\log(1/LC_{50})_{FM} = 0.834(0.17)\ \log(1/LC_{50})_{DM} + 0.327 \quad (3.70)$$

n=4 s=0.639 r^2adj=0.886 F=24.2

halogenated compounds:

$$\log(1/LC_{50})_{FM} = 0.807(0.074)\ \log(1/LC_{50})_{DM} + 0.177 \quad (3.71)$$

n=19 s=0.438 r^2adj=0.869 F=120.0

nitro compounds:

$$\log(1/LC_{50})_{FM} = 1.47(0.37)\ \log(1/LC_{50})_{DM} - 0.854 \quad (3.72)$$

n=4 s=0.367 r^2adj=0.830 F=15.6

aliphatic compounds:

$$\log(1/LC_{50})_{FM} = 0.855(0.10)\ \log(1/LC_{50})_{DM} - 0.028 \quad (3.73)$$

n=22 s=0.705 r^2adj=0.755 F=65.8

esters:

$$\log(1/LC_{50})_{FM} = 0.654(0.22)\ \log(1/LC_{50})_{DM} + 0.089 \quad (3.74)$$

n=4 s=1.065 r^2adj=0.727 F=9.0

Again there is a considerable variation in the slopes (1.47 (nitro compounds) to 0.654 (esters)) and the intercepts (0.327 (alkenes) to -0.854 (nitro compounds)), in the equations. The other, however less significant, relationship found was:

aromatic compounds:

$$\log(1/LC_{50})_{FM} = 0.658(0.11)\ \log(1/LC_{50})_{DM} + 0.342 \qquad (3.75)$$

n=24 s=0.640 r^2_{adj}=0.584 F=33.3

3.3.2.2. Analysis of the Outliers from the Relationship

Of the 46 chemicals, 12 (26%) were found to be significant outliers as shown in Table 3.20. With few outliers (seven relatively more toxic to the fathead minnow, and five to D. magna) it is difficult to distinguish any pattern; however, there are four compounds (three alkyl halides, and an aldehyde) that may act by attacking nucleophilic sites, that are more toxic to fathead minnow.

3.3.2.3. Chi Squared Analysis of Outliers

Table 3.21 summarises the results of the chi-squared analysis. Unfortunately, due to the lack of data, only the three calculations shown could be performed. None of these structural features shows a significant deviation from the expected ratio, although the analysis indicates that halogenated compounds are more toxic to the fathead minnow, and none is more toxic to D. magna, whilst alcohols tend to be more toxic to D. magna and not to the fathead minnow.

3.3.2.4. Analysis of the Outliers from the Toxicity-Hydrophobicity Relationship

The results of the chi-squared analysis of the outliers from the fathead minnow toxicity-ClogP (see Fig 3.6) and D. magna toxicity-ClogP (see Fig 3.9) relationships are summarised in Table 3.22. Although the chi-squared analysis shows there is no significant deviation from the null hypothesis for the outliers from the toxicity-ClogP relationships, four of the five chemicals more toxic than expected to D. magna, are outliers from the ClogP-D. magna toxicity relationship, as opposed to only one of these outliers being an outlier to the fathead minnow toxicity-ClogP relationship.

3.3.2.5. Improvement of the Relationship using Additional Parameters

A significant improvement is obtained in the interspecies relationship when the following parameters are included

$$\log(1/LC_{50})_{FM} = 0.598(0.076)\ \log(1/LC_{50})_{DM} + 0.298(0.070)\ ClogP - 0.445 \qquad (3.76)$$

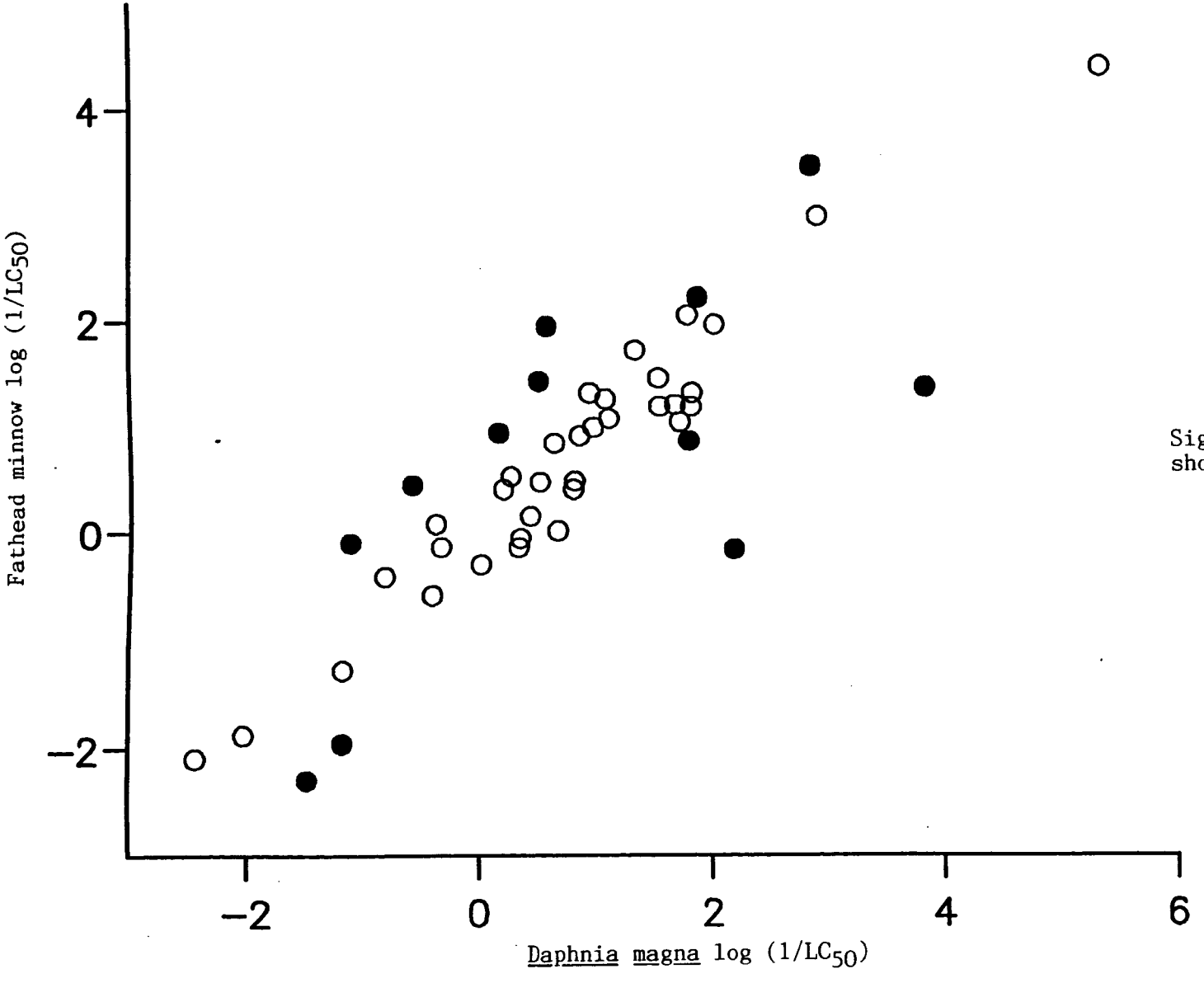
n=46 s=0.572 r^2_{adj}=0.820 F=103.6

$$\log(1/LC_{50})_{FM} = 0.693(0.082)\ \log(1/LC_{50})_{DM} + 0.338(0.068)\ ClogP - 0.166(0.068)\ lPV - 0.104 \quad (3.77)$$

$n=46 \qquad s=0.542 \qquad r^2_{adj}=0.839 \qquad F=78.9$

where lPV is the first order valence-corrected path molecular connectivity. These results were confirmed using best-subsets regression analysis. Equations 3.76 and 3.77 show, yet again, the importance of hydrophobicity, probably in the form of the chemicals' ability to be transported in the organism. Also obviously important is the bulk of the molecule (modelled by the molecular connectivity term) which again may be expressing some factor of its transport within the different species. There is an acceptable correlation between ClogP and lPV, $r=0.564$.

Figure 3.8 Graph of fathead minnow toxicity against Daphnia magna toxicity.



Significant outliers are shown as solid circles.

Figure 3.9 Graph of Daphnia magna toxicity against ClogP, showing outliers.

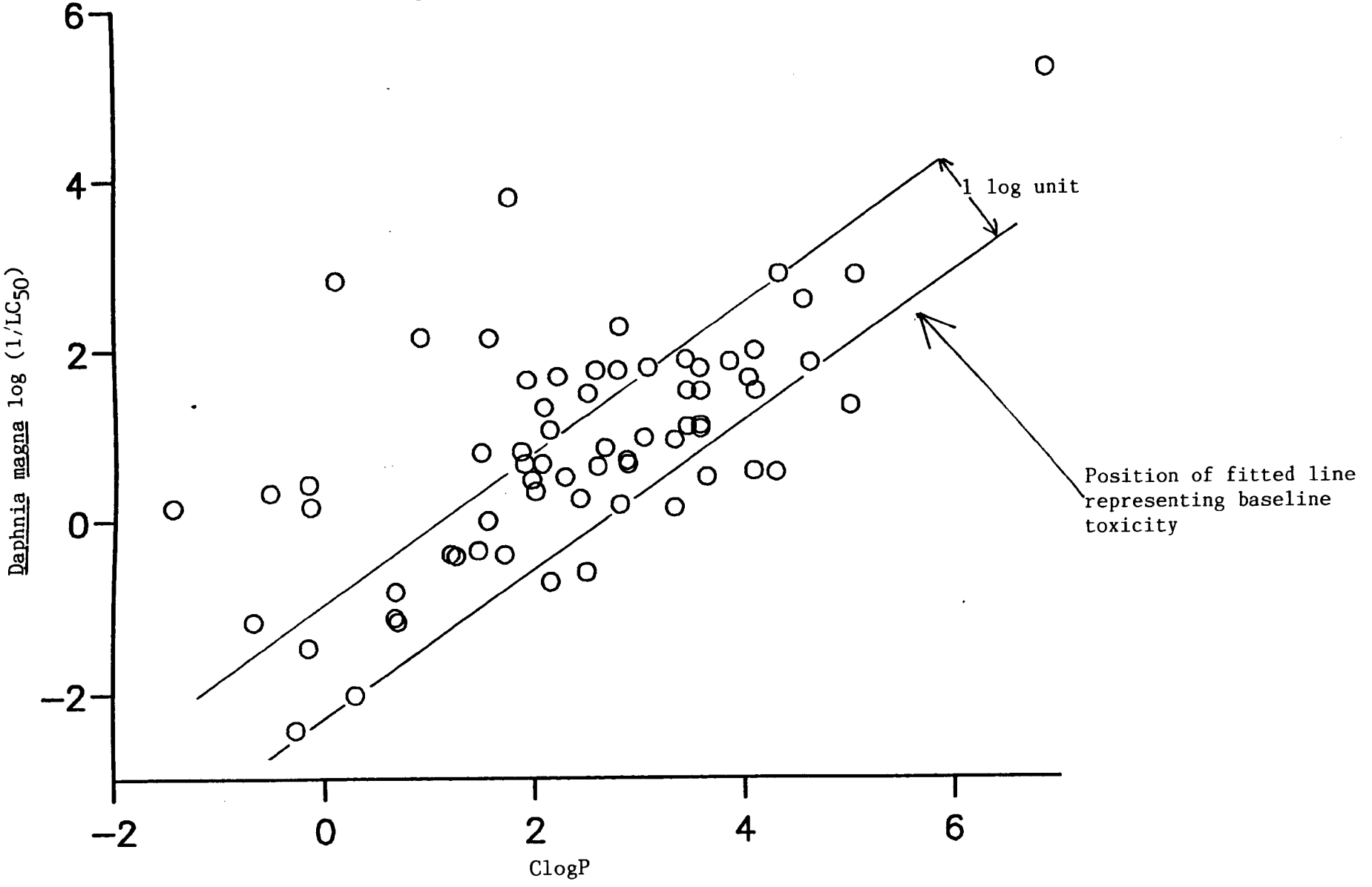


Table 3.20 Classification of Compounds in the Fathead Minnow and Daphnia magna Relationship

Chemical	Ratio Calc/Obs	Clogp Outlier*		Mode*
Chemicals More Toxic to Fathead Minnow				
acenaphthalene	0.03			
acrolein	0.076	F	D	4
pentachloroethane	0.11			4
1,1,1-trichloroethane	0.13			4
ethylbenzene	0.17			
pyridine	0.18	F		
hexachloroethane	0.21			4
Non-Outliers				
pentachlorophenol	0.25			
salicylaldehyde	0.25	F	D	
4,6-dinitro-2-cresol	0.27	F	D	
naphthalene	0.31			
1,3-dichlorobenzene	0.46			
1,3-dichloropropane	0.46			
dibenzofuran	0.50			
diethyl phthalate	0.53			
heptanol	0.54			
toluene	0.64			
ethyl acetate	0.65	F		
2,4,6-trichlorophenol	0.66			
1,1,2,2-tetrachloroethane	0.67			
tetrachloroethylene	0.70			
4-xylene	0.73			
1,2-dichloroethane	0.83			
permethrin	0.87			
trichloroethylene	0.98			
1,2,4-trimethylbenzene	1.24			
2,4-dinitrophenol	1.52	F	D	
2,4-dichlorophenol	1.53		D	
acetone	1.58			
4-nitrophenol	1.67		D	
2-chloroethanol	1.77	F	D	
phenol	1.96		D	
1,2-dichlorobenzene	2.03			
propanol	2.06			
dichloromethane	2.08			
2,2,2-trichloroethanol	2.30			
2-chlorophenol	2.38		D	
1,2-dichloropropane	2.45			
2-methylpropanol	2.51			
2,4-pentanedione	2.91	F	D	
nitrobenzene	3.77		D	
Chemicals More Toxic to Daphnia magna				
2,4-dimethylphenol	4.12		D	5
2-methyl-2,4-pentanediol	11.78		D	
2-(2-ethoxyethoxy)ethanol	14.79			
malathion	56.58	F	D	

aniline	90.74	D	3

* refer to Table 3.17 for full explanation

Table 3.21 Chi-Squared Analysis on the Outliers from the Fathead Minnow and *Daphnia magna* Relationship by Structural Features of the Chemicals

	More toxic to fathead minnow	Non-outlier	More toxic to *D. magna*	Chi-squared value
Overall relationship				
	7	34	5	
Aromatic compounds				
Observed	3	19	2	
Expected	3.5	18	2.5	0.226
Halogenated compounds				
Observed	3	16	0	
Expected	3	14	2	2.29
Alcohols				
Observed	0	14	3	
Expected	2.5	12.5	2	3.18

Table 3.22 Chi-Squared Analysis on the Outliers from the Toxicity-ClogP Relationship

		More toxic to fathead minnow	non-outliers	More toxic to *D. magna*	TOTAL
Fathead minnow-ClogP Outliers	O:	2	6	1	9
	E:	1.1	6.0	1.9	
Daphnia-ClogP Outliers	O:	1	10	4	15
	E:	1.9	10.0	3.1	
Total		3	16	5	24
Chi-squared		0.681	0.000	0.408	
		0.408	0.000	0.245	= 1.742

3.3.3.1. Inter-Species Relationship Between Fathead Minnow and Tetrahymena pyriformis

74 corresponding data were found for fathead minnow and T. pyriformis toxicity, and these show a very strong relationship (see Fig 3.10):

$$\log(1/LC_{50})_{FM} = 0.990(0.056)\ \log(1/IG_{50})_{TP} + 0.352 \quad (3.78)$$

n=74 s=0.443 r^2_{adj}=0.807 F=307.2

where: $\log(1/IG_{50})_{TP}$ is the log (1/48-60 hour IG_{50}) to the ciliate Tetrahymena pyriformis.

When the data were considered according to structural features, even better correlations are achieved for some sub-sets:

ethers:

$$\log(1/LC_{50})_{FM} = 1.13(0.099)\ \log(1/IG_{50})_{TP} + 0.263 \quad (3.79)$$

n=7 s=0.226 r^2_{adj}=0.956 F=130.3

ketones:

$$\log(1/LC_{50})_{FM} = 1.07(0.20)\ \log(1/IG_{50})_{TP} - 0.162 \quad (3.80)$$

n=4 s=0.230 r^2_{adj}=0.905 F=29.6

nitrogen ring compounds:

$$\log(1/LC_{50})_{FM} = 1.03(0.12)\ \log(1/IG_{50})_{TP} + 0.372 \quad (3.81)$$

n=15 s=0.363 r^2_{adj}=0.848 F=78.9

alcohols:

$$\log(1/LC_{50})_{FM} = 0.949(0.082)\ \log(1/IG_{50})_{TP} + 0.425 \quad (3.82)$$

n=31 s=0.405 r^2_{adj}=0.816 F=134.1

Other structural features that produced weaker correlations included the

halogenated compounds:

$$\log(1/LC_{50})_{FM} = 0.981(0.13)\ \log(1/IG_{50})_{TP} + 0.333 \quad (3.83)$$

n=24 s=0.551 r^2_{adj}=0.697 F=54.0

amines:

$$\log(1/LC_{50})_{FM} = 1.05(0.15)\ \log(1/IG_{50})_{TP} + 0.153 \quad (3.84)$$

n=24 s=0.527 r^2_{adj}=0.689 F=51.9

nitro compounds:

$$\log(1/LC_{50})_{FM} = 0.860(0.191)\ \log(1/IG_{50})_{TP} + 0.350 \qquad (3.85)$$

n=19 s=0.462 r^2_{adj}=0.516 F=20.2

3.3.3.2. Analysis of the Outliers from the Inter-Species Relationship

Of the 74 compounds present in the relationship only 11 (or 15%) were classified as outliers. The outliers from the inter-species relationship are listed in Table 3.23.

3.3.3.3. Chi-Squared Analysis on Structural Features of the Outliers

The chi-squared analysis (summarised in Table 3.24) shows that none of the structural features tested was significant at the 95% level. However, compounds containing an amine group (in this case the anilines) were more prone than might be expected to be outliers.

3.3.3.4. Analysis of the Outliers from the Toxicity-Hydrophobicity Relationship

When outliers from the fathead minnow toxicity-ClogP and T. pyriformis toxicity-ClogP (see Fig 3.11) relationships were considered (see Table 3.25), although again there is no significant deviation from the expected ratio, more outliers from the fathead minnow toxicity-ClogP relationship (than from T. pyriformis) are relatively more toxic to fathead minnow, and conversely more outliers from T. pyriformis toxicity-ClogP relationship are relatively more toxic to T. pyriformis.

3.3.3.5. Improvement of the Relationship using Additional Parameters

After stepwise regression analysis using the parameters described in section 2.4.6.1, a slight improvement is made in the relationship:

$$\log(1/LC_{50})_{FM} = 0.637(0.086)\ \log(1/IG_{50})_{TP} + 0.335(0.065)\ \text{Clogp} - 0.248 \qquad (3.86)$$

n=72 s=0.384 r^2_{adj}=0.850 F=203

$$\log(1/LC_{50})_{FM} = 0.591(0.087)\ \log(1/IG_{50})_{TP} + 0.280(0.069)\ \text{ClogP} + 0.149(0.074)\ 3PV - 0.322 \qquad (3.87)$$

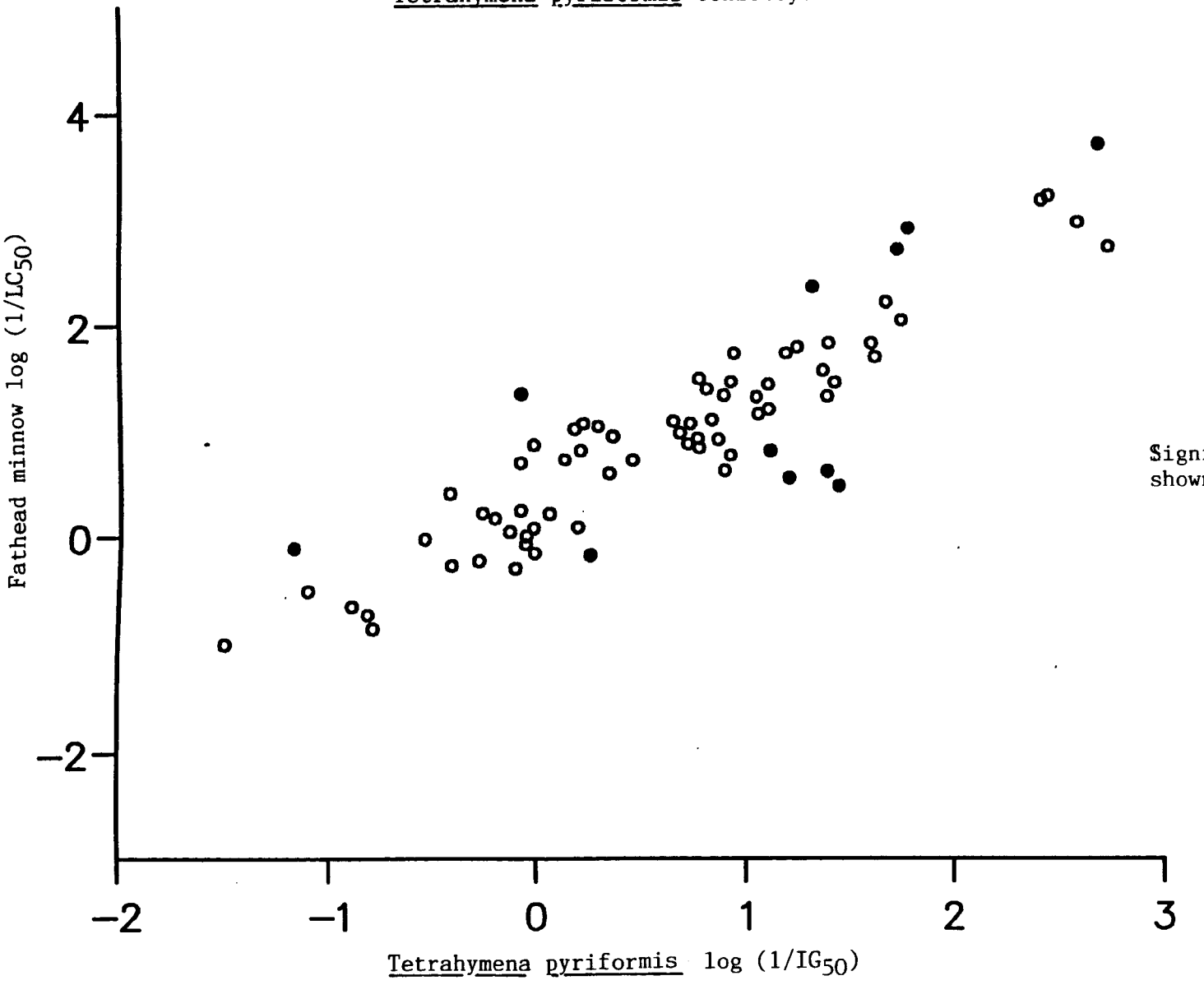
n=72 s=0.376 r^2_{adj}=0.857 F=142

where: 3PV is the third order valence corrected molecular connectivity.

Again ClogP is important, representing the transport of a chemical into the

organism; however, it adds little to the value of the equation due to the already strong correlation present. In addition, due to the very small increase in r^2, the third (bulk) parameter cannot be assumed to be making any significant contribution to the relationship. These equations were confirmed using best-subsets regression analysis.

Figure 3.10 Graph of fathead minnow toxicity against Tetrahymena pyriformis toxicity.



Significant outliers are shown as solid circles.

Figure 3.11 Graph of Tetrahymena pyriformis toxicity against ClogP, showing outliers.

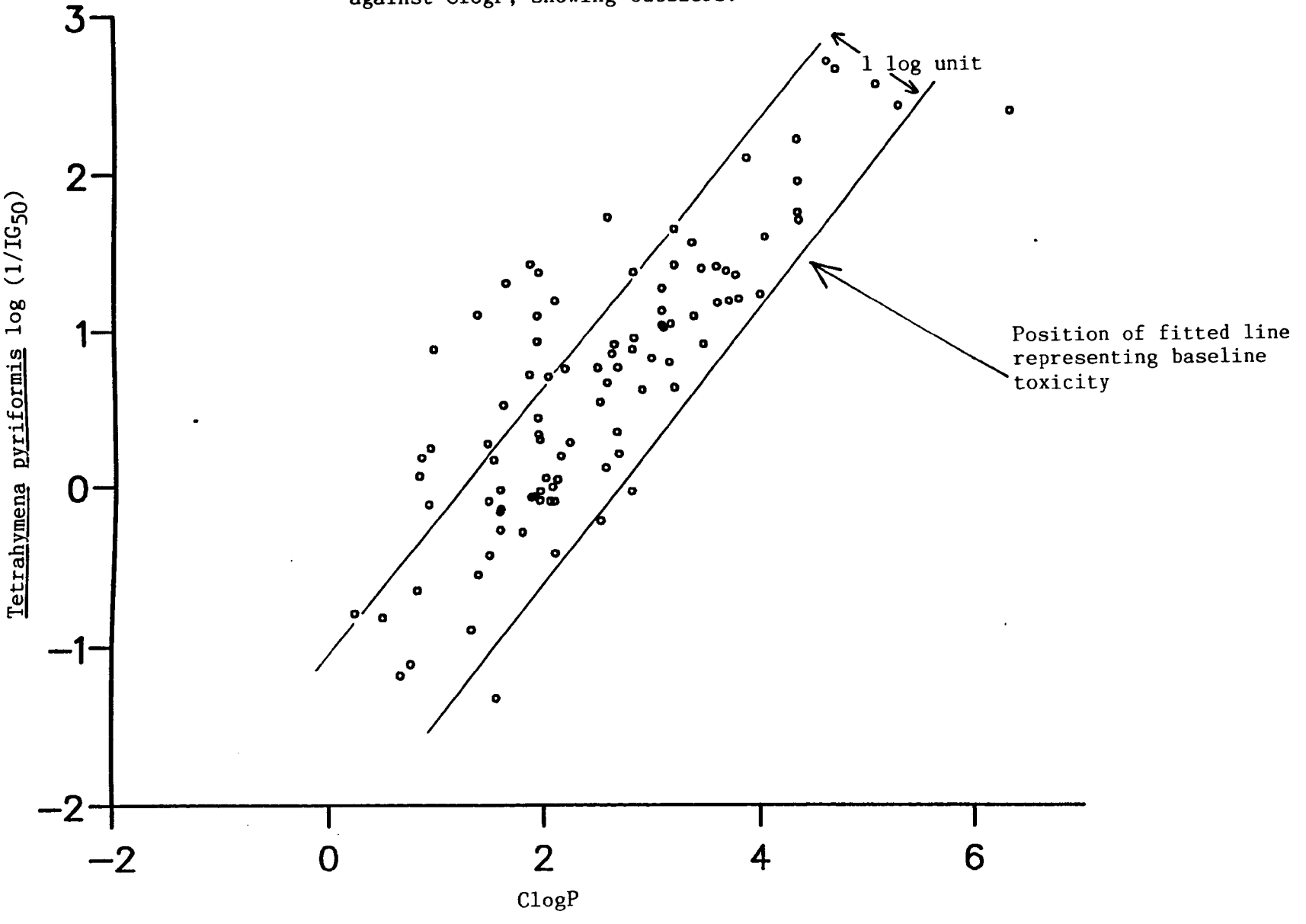


Table 3.23 Classification of the Compounds in the Fathead Minnow and Tetrahymena pyriformis Relationship

Chemical	Ratio Calc/Obs	ClogP Outlier*		Mode*
Chemicals More Toxic to Fathead Minnow				
2-chloroaniline	0.08	F		
2,3,5,6-tetrachloroaniline	0.15			2
1,4-dinitrobenzene	0.19	F		2
pyridine	0.19	F		
pentabromophenol	0.19	F		2
pentachloropyridine	0.21			
Non-Outliers				
2,4-dimethylphenol	0.29			
4-ethylphenol	0.31			
2-nitrobenzaldehyde	0.31	F		
4-octylaniline	0.33			
phenol	0.33			
2,5-dinitrophenol	0.34	F	T	
4-nonylphenol	0.34			
2-chlorophenol	0.38			
4-fluoronitrobenzene	0.38			
1-naphthol	0.41			
4-methylphenol	0.53			
3-nitrotoluene	0.55			
4-chloro-3-methylphenol	0.55			
2-allylphenol	0.55			
pyrrole	0.57			
4-phenylazophenol	0.58	F		
4-tert-pentyl phenol	0.59			
2,3,4-trichloroaniline	0.60			
4-tert-butylphenol	0.62			
2-dimethylaminopyridine	0.66			
1-benzylpyridinum 3-sulphonate	0.71			
3-methoxyphenol	0.72			
4-hexyloxyaniline	0.76			
3-benzyloxyaniline	0.76			
4-propylphenol	0.77			
pentachlorophenol	0.81			
5-ethyl-2-methylpyridine	0.92			
2,4-dinitroaniline	0.97	F	T	
2-phenylphenol	0.98			
4,6-dinitro-2-cresol	1.02	F	T	
4-benzoylpyridine	1.03			
4-phenylpyridine	1.06			
4-bromophenyl-3-pyridylketone	1.13			
2,4-dichlorophenol	1.13			
2-amino-5-chlorobenzonitrile	1.15			
5-hydroxy-2-nitrobenzaldehyde	1.19			
2,4-dinitro-1-naphthol, sodium salt	1.21			
4-picoline	1.26			
4-phenoxyphenol	1.31			
4-methoxyphenol	1.44			
2-chloro-4-nitroaniline	1.46			
methyl 4-nitrobenzoate	1.47		T	

4-ethylaniline	1.50			
6-chloro-2-picoline	1.58			
4-butylaniline	1.66			
2,4-dinitrophenol	1.69	F	T	
2,4,6-tribromophenol	1.71			
1,2-bis(4-pyridyl)ethane	1.72			
4-acetamidophenol	1.80			
4-ethoxy-2-nitroaniline	1.81			
3-chloronitrobenzene	1.86			
nitrobenzene	1.88			
2,3,4,5-tetrachlorophenol	1.92			
6-chloro-2-pyridinol	1.94			
2,4,6-trichlorophenol	1.94			
2-nitrophenol	2.22			
3,4-dichloroaniline	2.41			
2-cyanopyridine	2.57		T	
4-nitrobenzamide	2.71	F	T	
4-toluidine	2.97			
a,a,a-4-tetrafluoro-3-toluidine	3.00			
2,6-pyridinedicarboxylic acid	3.36		T	
4-amino-2-nitrophenol	3.93	F	T	
Chemicals More Toxic to Tetrahymena				
4-fluoroaniline	4.22	F	T	3
aniline	5.61		T	3
4-bromoaniline	9.38		T	3
4-chloroaniline	12.31		T	3
4-nitrophenol	19.07		T	3

* refer to Table 3.17 for full explanation

Table 3.24 Chi-Squared Analysis on the Outliers from the Fathead Minnow and Tetrahymena pyriformis Relationship by the Structural Features of the Chemicals

	More toxic to fathead minnow	Non-outlier	More toxic to T. pyriformis	Chi-squared value
Overall relationship				
	6	63	5	
halogenated compounds				
Observed	4	17	3	
Expected	2	20	1.5	3.95
alcohols				
Observed	1	29	1	
Expected	2.5	26.5	2	1.64
amines				
Observed	2	16	4	
Expected	1.5	19	1.5	4.81
nitrogen ring compounds				
Observed	2	15	0	
Expected	1.5	14.5	1	1.19
nitro compounds				
Observed	1	16	1	
Expected	1.5	15	1.5	0.41

Table 3.25 Chi-Squared Analysis on the Outliers from the Toxicity-ClogP Relationship

		More toxic to fathead minnow	non-outliers	More toxic to Tetrahymena	TOTAL
Fathead minnow-ClogP Outliers	O:	4	8	1	13
	E:	1.9	8.2	2.9	
Tetrahymena-ClogP Outliers	O:	0	9	5	14
	E:	2.1	8.8	3.1	
Total		4	17	6	27
Chi-squared		2.234	0.004	1.235	
		2.074	0.004	1.147	= 6.70

3.3.4.1. Inter-Species Relationship Between Daphnia magna and Microtox

Overall 52 corresponding toxicity data were obtained for the D. magna and Microtox tests. These gave the following relationship (see Fig 3.12):

$$\log(1/LC_{50})_{DM} = 0.782(0.085)\ \log(1/EC_{50})_{M} + 0.295 \quad (3.88)$$

n=52 s=0.891 r^2adj=0.619 F=83.8

When the data were analysed according to the structural features they contain, some extremely good relationships were observed:

esters:

$$\log(1/LC_{50})_{DM} = 1.53(0.073)\ \log(1/EC_{50})_{M} + 1.05 \quad (3.89)$$

n=4 s=0.252 r^2adj=0.993 F=434.8

nitro compounds:

$$\log(1/LC_{50})_{DM} = 1.43(0.383)\ \log(1/EC_{50})_{M} - 0.386 \quad (3.90)$$

n=4 s=0.247 r^2adj=0.812 F=14.0

alkenes:

$$\log(1/LC_{50})_{DM} = 1.71(0.37)\ \log(1/EC_{50})_{M} - 0.492 \quad (3.91)$$

n=6 s=0.892 r^2adj=0.798 F=20.7

Other weaker relationships are found:

alcohols:

$$\log(1/LC_{50})_{DM} = 0.801(0.096)\ \log(1/EC_{50})_{M} + 0.405 \quad (3.92)$$

n=21 s=0.649 r^2adj=0.773 F=69.2

aliphatic compounds:

$$\log(1/LC_{50})_{DM} = 0.705(0.13)\ \log(1/EC_{50})_{M} + 0.200 \quad (3.93)$$

n=22 s=0.942 r^2adj=0.577 F=29.7

aromatic compounds:

$$\log(1/LC_{50})_{DM} = 0.832(0.17)\ \log(1/EC_{50})_{M} + 0.286 \quad (3.94)$$

n=30 s=0.872 r^2adj=0.438 F=23.6

halogenated compounds:

$$\log(1/LC_{50})_{DM} = 0.609(0.15)\ \log(1/EC_{50})_{M} + 0.595 \quad (3.95)$$

n=24 s=0.959 r^2adj=0.399 F=16.3

amines:

$$\log(1/LC_{50})_{DM} = 0.620(1.04)\ \log(1/EC_{50})_{M} + 0.854 \qquad (3.96)$$

n=4 s=1.347 r^2_{adj}=0.000 F=0.36

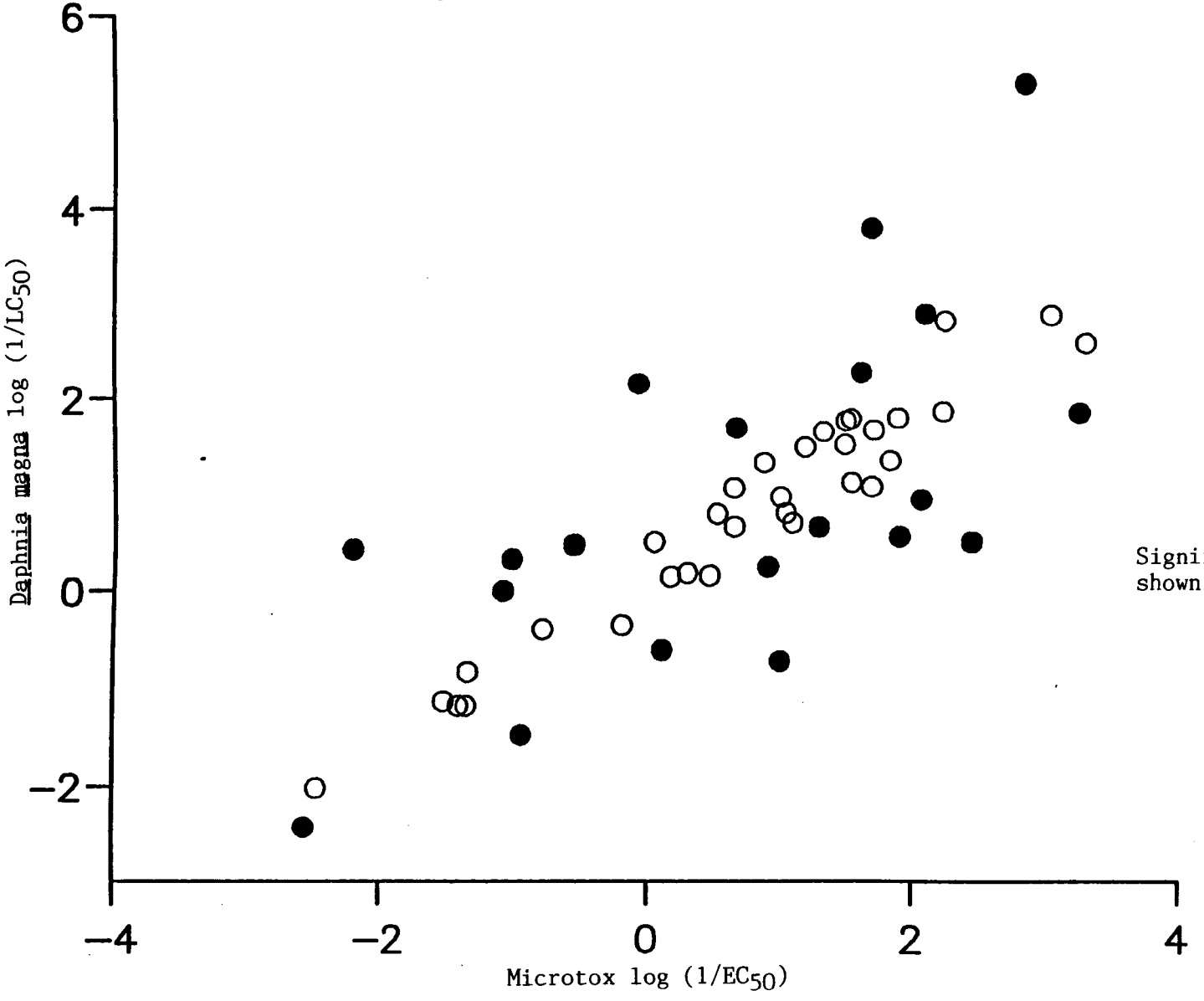
3.3.4.2. Analysis of the Outliers from the Inter-Species Relationship

Of the 52 data in the overall relationship, 20 (38%) were found to be significant outliers. The outliers from the inter-species relationships are listed in Table 3.26.

3.3.4.3. Chi-Squared Analysis on Structural Features

The summary of the chi-squared analysis on the structural features present in the molecules is shown in Table 3.27. It shows that there are only four groups with enough data to enable the analysis to be performed; none of these chemical groups appears significantly to affect the toxicity of the chemicals to either of the species.

Figure 3.12 Graph of Daphnia magna toxicity against Microtox toxicity.



Significant outliers are shown as solid circles.

Table 3.26 Classification of Compounds in the Daphnia magna and Microtox Relationship

Chemical	Ratio Calc/Obs	Mode*
Chemicals More Toxic to Daphnia magna		
permethrin	0.0016	
malathion	0.0065	
aniline	0.012	3
2-chloroethanol	0.014	
2,3,4,6-tetrachlorophenol	0.11	2
2-chlorophenol	0.13	
2,4-pentanedione	0.15	
acrolein	0.16	4
2,4-dichloroaniline	0.19	3
chloroform	0.24	
Non-outliers		
2,2,2-trichloroethanol	0.28	
salicyladehyde	0.45	
2,4-dinitrophenol	0.47	
4,6-dinitro-2-cresol	0.50	
1,2-dichlorobenzene	0.51	
4-chlorophenol	0.53	
2-cresol	0.55	
pentachlorophenol	0.59	
trichloroethylene	0.65	
phenol	0.80	
2,4,6-trichlorophenol	0.86	
biphenyl	0.90	
2,4-dichlorophenol	0.93	
ethyl acetate	1.18	
ethylpropioate	1.19	
tetrachloroethylene	1.27	
nitrobenzene	1.37	
2,4,5-trichlorophenol	1.47	
pyridine	1.72	
dibenzothiophene	1.83	
diethanolamine	1.87	
4-nitrophenol	2.00	
toluene	2.16	
2-methyl-2,4-pentanediol	2.34	
1,2,3,5-tetrachlorobenzene	2.35	
1,4-dichlorobenzene	2.38	
propanol	2.42	
2-methylpropanol	2.60	
chlorobenzene	2.75	
1,2-dichloroethane	3.06	
allyl amine	3.12	
1,3-dichlorobenzene	3.46	
More toxic to Microtox		
styrene	4.36	
acetone	5.28	
heptanol	5.58	

naphthalene	9.12	
hexachloroethane	9.35	4
1,1,1-trichloroethane	9.38	4
2-(2-ethoxyethoxy)ethanol	11.01	
1,2,4-trichlorobenzene	16.33	
pentachloroethane	48.89	4
benzene	61.09	

* refer to Table 3.17 for full explanation

Table 3.27 Chi-Squared Analysis on the Outliers from the Daphnia magna and Microtox Relationship by Structural Features of the Chemicals

	More toxic to fathead minnow	Non-outlier	More toxic to Microtox	Chi-squared value
Overall relationship				
	10	32	10	
Aromatic compounds				
Observed	5	21	4	
Expected	6	18	6	1.34
Halogenated compounds				
Observed	6	13	5	
Expected	4.5	15	4.5	0.82
Alcohols				
Observed	3	16	2	
Expected	4	13	4	1.94
Alkenes				
Observed	2	3	1	
Expected	1	4	1	1.25

3.3.5.1. Inter-Species Relationship Between Daphnia magna and Tetrahymena pyriformis

17 data were found for the corresponding toxicities, and these gave only a weak relationship (see Fig 3.13):

$$\log(1/LC_{50})_{DM} = 0.577(0.17)\ \log(1/IG_{50})_{TP} + 1.03 \qquad (3.97)$$

n=17 s=0.699 r^2adj=0.381 F=10.8

When the data were analysed according to the structural features they contain, the following correlations were found:

halogenated compounds

$$\log(1/LC_{50})_{DM} = 0.441(0.19)\ \log(1/IG_{50})_{TP} + 1.29 \qquad (3.98)$$

n=6 s=0.381 r^2adj=0.462 F=5.3

alcohols:

$$\log(1/LC_{50})_{DM} = 0.361(0.16)\ \log(1/IG_{50})_{TP} + 1.25 \qquad (3.99)$$

n=11 s=0.477 r^2adj=0.275 F=4.8

nitro compounds:

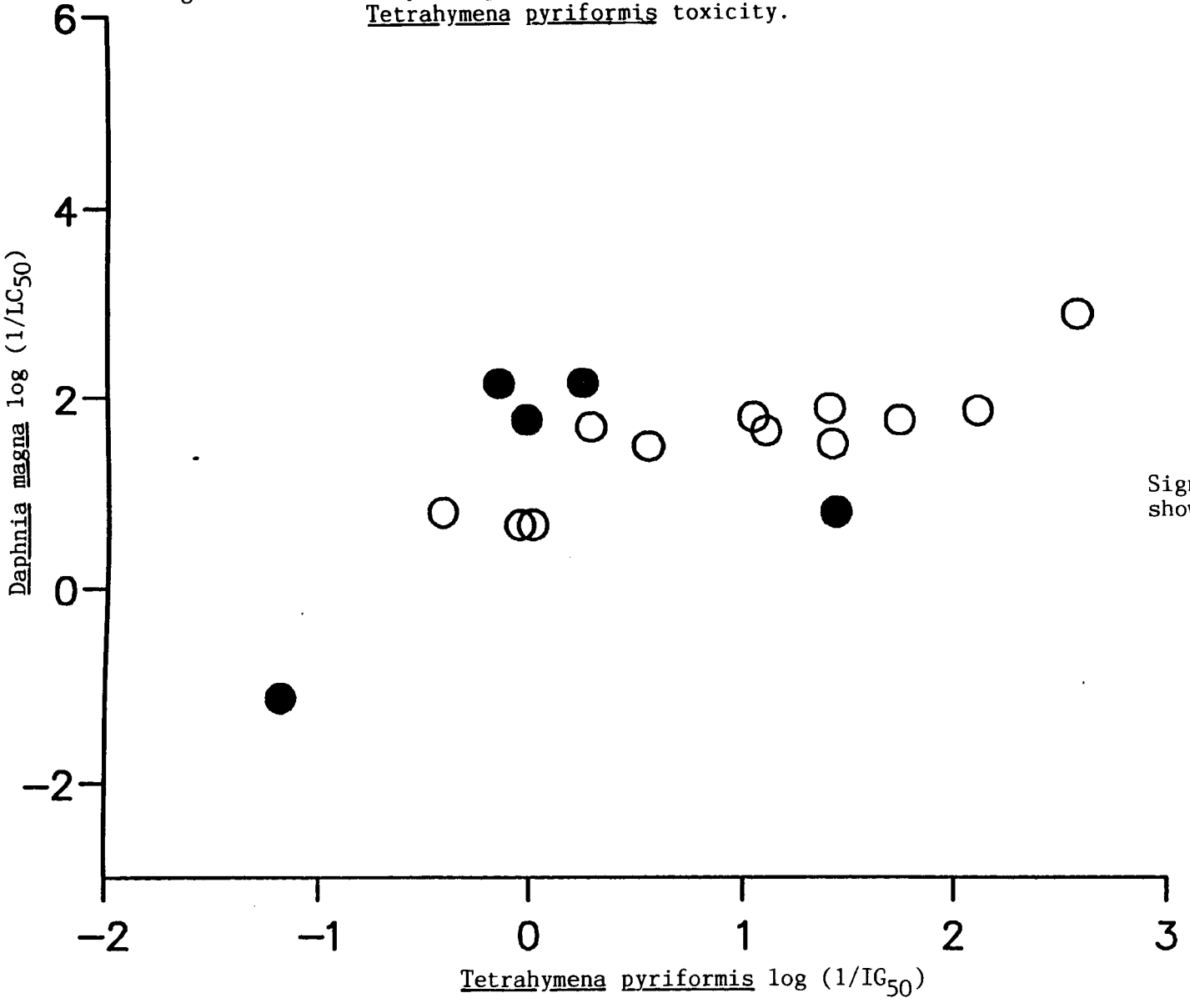
$$\log(1/LC_{50})_{DM} = 0.464(0.40)\ \log(1/IG_{50})_{TP} + 0.733 \qquad (3.100)$$

n=4 s=0.537 r^2adj=0.112 F=1.38

3.3.5.2. Analysis of the Outliers from the Relationship

Of the data in the overall inter-species relationship, five (29%) are significant outliers, and these are listed in Table 3.28. With such a small number it would presumptuous to draw firm conclusions from the data concerning mode of action and which structural features may be of importance.

Figure 3.13 Graph of Daphnia magna toxicity against Tetrahymena pyriformis toxicity.



Significant outliers are shown as solid circles.

Table 3.28 Classification of Compounds in the Daphnia magna and Tetrahymena pyriformis Relationship

Chemical	Ratio Calc/Obs	Mode*
Chemicals More Toxic to Daphnia magna		
3-methylaniline	0.061	
aniline	0.10	
2,4-dimethylphenol	0.18	5
Non-outliers		
2-chlorophenol	0.32	
pentachlorophenol	0.42	
2,4-dichlorophenol	0.68	
4-chlorophenol	0.71	
acridine	0.88	
phenol	0.98	
2,4-dinitrophenol	1.02	
4,6-dinitro-2-cresol	1.83	
2,4,6-trichlorophenol	2.12	
nitrobenzene	2.16	
quinoline	2.33	
2,4,5-trichlorophenol	2.39	
Chemicals More Toxic to Tetrahymena		
4-nitrophenol	11.23	3
pyridine	30.14	

* refer to Table 3.17 for full explanation

3.3.6.2. Inter-Species Relationship Between Microtox and Tetrahymena pyriformis

A total of 54 corresponding toxicity data were obtained giving the following relationship (see Fig 3.14):

$$\log(1/EC_{50})_M = 0.610(0.12)\ \log(1/IG_{50})_{TP} + 1.07 \quad (3.101)$$

n=54 s=0.746 r^2adj=0.317 F=25.5

Equation 3.101 was moderately improved by separating the compounds into chemical classes:

halogenated compounds:

$$\log(1/EC_{50})_M = 0.710(0.15)\ \log(1/IG_{50})_{TP} + 0.817 \quad (3.102)$$

n=23 s=0.518 r^2adj=0.479 F=21.2

alcohols:

$$\log(1/EC_{50})_M = 0.454(0.13)\ \log(1/IG_{50})_{TP} + 1.28 \quad (3.103)$$

n=36 s=0.674 r^2adj=0.244 F=12.3

amines:

$$\log(1/EC_{50})_M = 0.637(0.36)\ \log(1/IG_{50})_{TP} + 0.931 \quad (3.104)$$

n=14 s=0.852 r^2adj=0.140 F=3.1

nitro compounds:

$$\log(1/EC_{50})_M = 0.524(0.37)\ \log(1/IG_{50})_{TP} + 0.752 \quad (3.105)$$

n=8 s=0.602 r^2adj=0.125 F=2.0

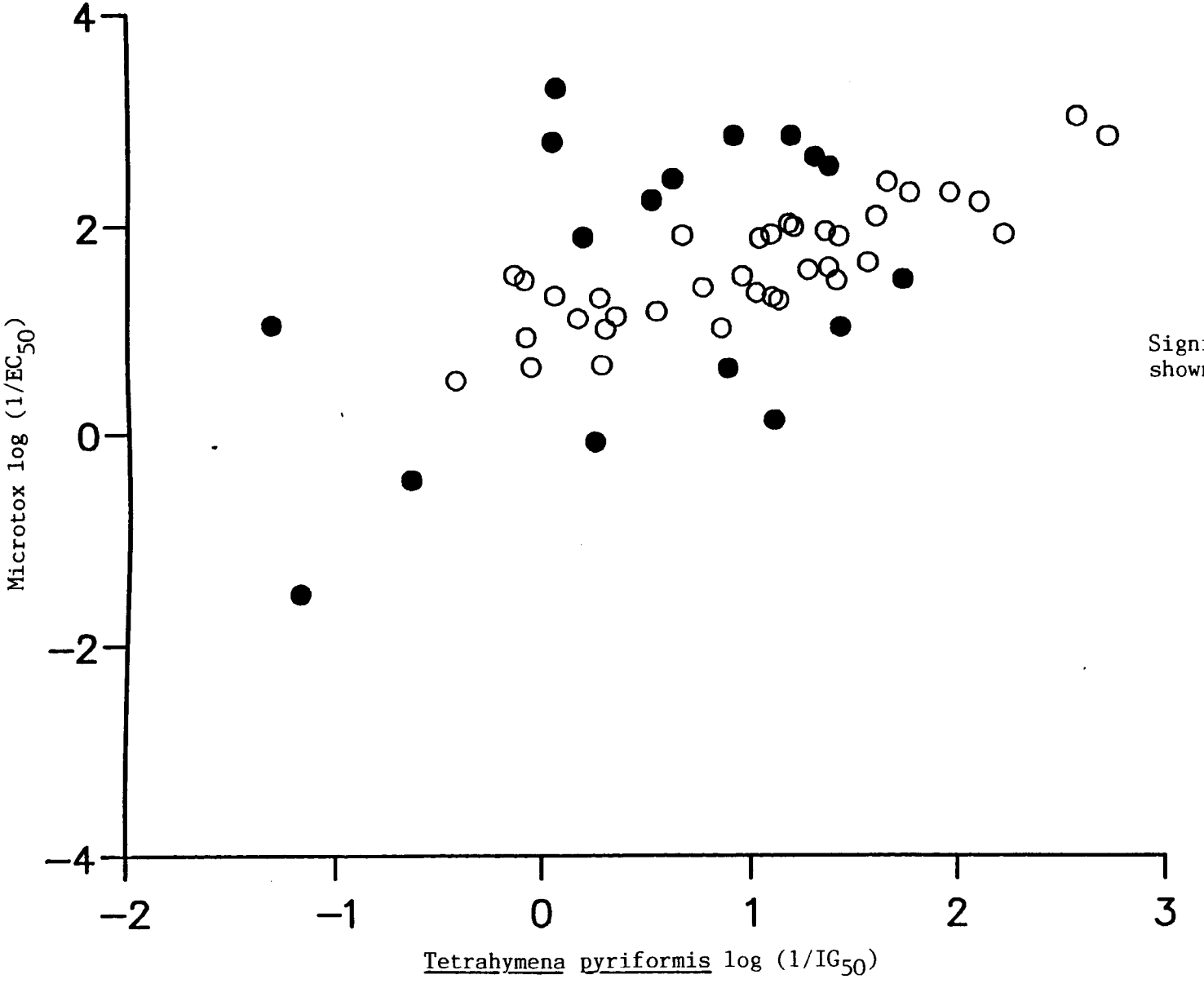
3.3.6.2. Analysis of the Outliers

Of the 54 data in the overall inter-species relationship, 17 (31%) were found to be significant outliers, and these are listed in Table 3.29.

3.3.6.3. Chi-Squared Analysis on Structural Features

There are enough data to analyse only three of the structural features (Table 3.30). None of these shows a significant deviation from the expected results. However, the higher chi-squared value for compounds containing halogens indicates that these compounds may be less likely to be outliers.

Figure 3.14 Graph of Microtox toxicity against Tetrahymena pyriformis toxicity.



Significant outliers are shown as solid cirles.

Table 3.29 Classification of the Compounds in the Microtox and Tetrahymena pyriformis Relationship

Chemical	Ratio Calc/Obs	Mode*
Chemicals More Toxic to Microtox		
hydroquinone	0.0064	
4-ethylaniline	0.020	3
4-tert-butylphenol	0.060	5
4-benzylphenol	0.088	
aaa-trifluoro-4-cresol	0.10	
4-cyanophenol	0.14	
4-hydroxybenzoic acid	0.16	
1,4-dinitrobenzene	0.16	3
4-methylphenol	0.20	5
3,4-dichloroaniline	0.22	3
Non-outliers		
4-methoxyphenol	0.28	
4-hydroxyacetophenone	0.34	
4-phenylpyridine	0.37	
pentachlorophenol	0.40	
4-phenylazophenol	0.47	
4-hydroxypropiophenone	0.59	
2,3,4-trichloroaniline	0.59	
4-chloro-3,5-dimethylphenol	0.65	
2,4-dichlorophenol	0.66	
2-phenylphenol	0.66	
2,3,5,6-tetrachloroaniline	0.68	
2,3,4,5-tetrachlorophenol	0.77	
4-hydroxybenzaldehyde	0.84	
4-phenoxyphenol	0.88	
2,3,4,5-tetrachloroaniline	0.90	
2,4,6-tribromophenol	0.91	
4-phenylazoaniline	1.09	
2-nitrobenzaldehyde	1.15	
2-chloroaniline	1.22	
1-naphthol	1.33	
2,4,5-trichlorophenol	1.35	
4-phenylaniline	1.35	
2-allylphenol	1.42	
4-chlorophenol	1.66	
3-chloroaniline	1.74	
2,3-dichlorophenol	1.84	
phenol	1.98	
4-chloroaniline	2.03	
4-hydroxybenzophenone	2.15	
3,5-dichlorophenol	2.36	
nitrobenzene	2.46	
2,4-dinitrophenol	2.62	
2,4,6-trichlorophenol	2.82	
2,5-dichlorophenol	2.94	
2,3,5,6-tetrachlorophenol	3.19	
3-chloronitrobenzene	3.70	
2-chlorophenol	3.79	

Chemicals More Toxic to Tetrahymena

4,6-dinitro-2-cresol	4.31	
4-nitrophenol	8.13	3
4-amino-2-nitrophenol	9.48	3
4-fluoroaniline	10.95	3
resorcinol	13.29	
aniline	19.82	3
pyridine	74.16	

* refer to Table 3.17 for full explanation

Table 3.30 Chi-Squared Analysis on the Outliers from the Microtox and Tetrahymena pyriformis Relationship by the Structural Features of the Chemicals

	More toxic to Microtox	Non-outlier	More toxic to T. pyriformis	Chi-squared value
Overall relationship				
	10	37	7	
Halogenated compounds				
Observed	1	19	1	
Expected	4	14	3	5.37
Alcohols				
Observed	7	25	4	
Expected	7	24.5	4.5	0.06
Amines				
Observed	2	9	3	
Expected	2.5	9.5	2	0.63

3.3.7. Results of the Investigation for the Most 'Representative' Test Species

The results in Table 3.31 indicate that the Tetrahymena assay has the lowest average positive residual when predicitng the toxicity of the other species (as described in section 2.4.6), of the four toxicity tests.

Table 3.31 The Sum of the Positive Residuals for the Regression Equations Involving the Aquatic Toxicity of the Following Species, the Number of Data Points, and the Average Positive Residual per Data Point

	Species used to predict toxicity			
	Fathead Minnow	Microtox	Daphnia magna	Tetrahymena pyriformis
Fathead Minnow	-	37.77 126	11.51 46	12.64 74
Microtox	44.78 126	-	17.38 52	14.58 54
Daphnia magna	11.74 46	16.57 52	-	4.21 17
Tetrahymena pyriformis	11.08 74	15.00 54	5.18 17	-
TOTALS	67.60 246	69.34 232	34.07 115	31.43 145
AVERAGE	0.275	0.299	0.296	0.217

The first number is the sum of the positive residuals in each relationship, and the second (lower) number is the number of data points.

4. DISCUSSION

4.1.1. Evaluation of Experimentally Determined Microtox Data

After relatively little experience, producing toxicity data using the Microtox analyser was found to be a rapid and simple task. The Microtox system was sensitive to a range of toxicity values from 21100 mg/l (for acetone) to 0.62 mg/l (for diethyl sebacate). The data were also reproducible when different stock solutions and samples of bacteria were applied. The only modification of the method needed was the application of methanol to aid the dissolution of less soluble organic chemicals.

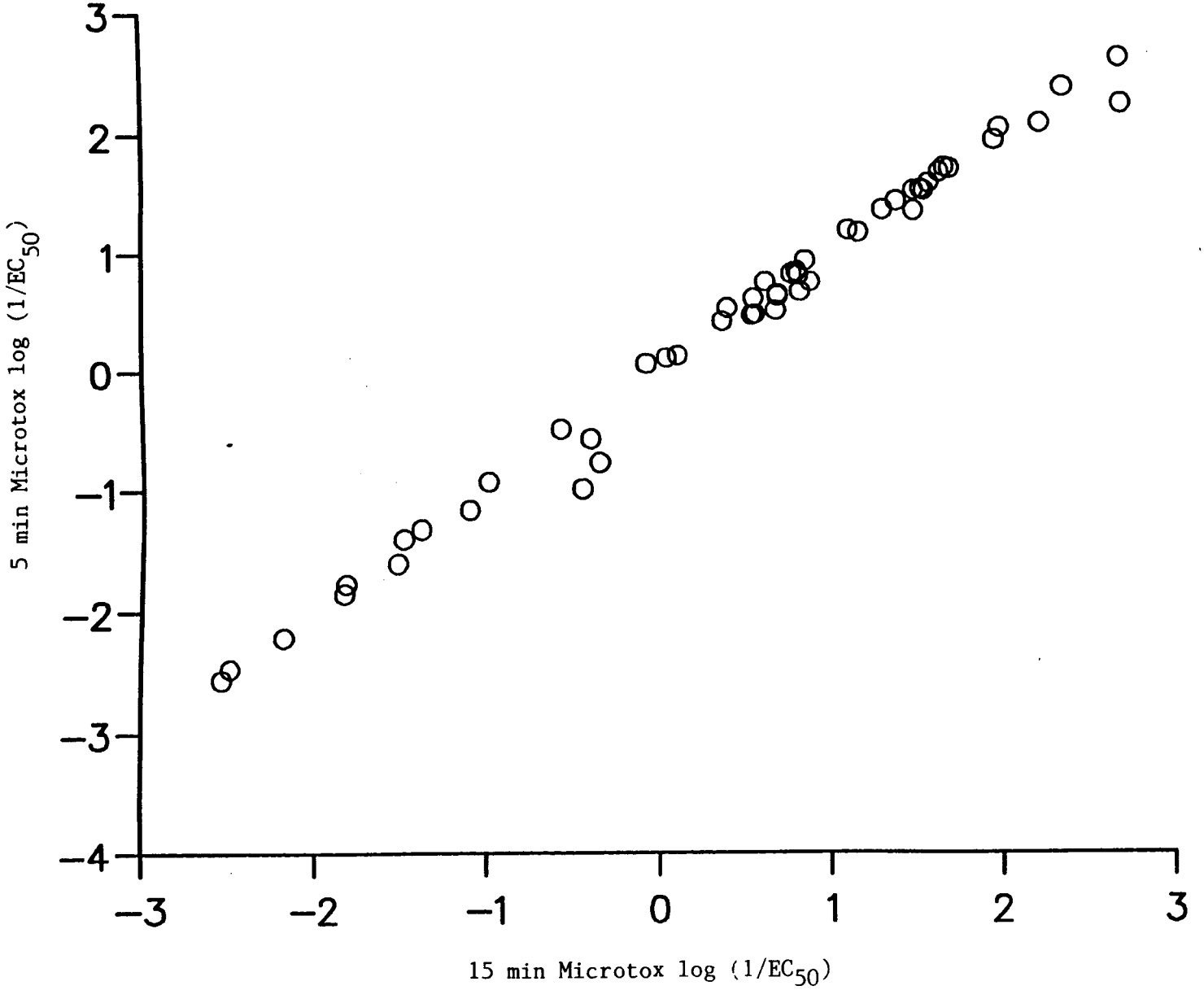
The 5 minute toxicity data are used throughout this study, although 15 minute data were also measured. 5 and 15 minute EC_{50}s from this study were found to be very well correlated. The relationship is shown in Fig 4.1 and expressed as:

$$5 \text{ min } \log(1/EC_{50})_M = 1.00(0.016)\ 15 \text{ min } \log(1/EC_{50})_M - 0.018 \qquad (4.1)$$

$n=48 \qquad s=0.140 \qquad r^2(adj)=0.988 \qquad F=3792$

The choice of 5 minute EC_{50} as the biological response in the QSAR (and interspecies) studies was simply because many more published data were available for this endpoint. There are, however, various advantages and disadvantages in using longer test times (e.g. 15 or 30 minutes) to find the EC_{50}. Some chemicals need a longer test time to interact with the target organism (Ribo and Kaiser, 1987), so a short test time may lead to misleading results. As well a lengthening of the test time may enable other toxic effects to take place, such as metabolism of the chemical into a more reactive form. One study at least, though (Beckman Inc, 1981), suggests that the toxic effect will have occured after 5 minutes for most organic chemicals, and greater toxicity at longer test times is important only for inorganic chemicals (e.g. metal complexes, commonly found in sludge samples, but not relevant in this study). Also the high correlation between the toxicities at the two test times in this study indicates that the toxic action has taken place after 5 minutes. A further disadvantage with longer exposure times is the decay in quality, and thus, light output of the bacteria, which may mean fewer 'runs' can be taken from one sample, thus possibly increasing the cost

Figure 4.1 Graph of 5 minute Microtox toxicity against 15 minute Microtox toxicity.



of the tests.

4.1.2. Comparison of Toxicity Data Obtained in this Study with Published Data

Of the 48 chemicals tested, published Microtox data were found for seven. These are summarised in Table 4.1. An extremely good agreement between the published toxicities of acetone and the four substituted benzenes and those obtained in this work can be seen. The toxicity of toluene obtained here is slightly higher than the published values (the value of -2.56 is very unusual, and should be ignored). The largest difference in toxicities is for butanal. The cause of this is open to conjecture, but it should be noted that the value obtained here was the mean of four replicates (standard error 0.045), whereas Curtis et al (1982) took only one measurement. In addition the value obtained in this study was well fitted in the relationship with ClogP (eqn 4.3), as would be expected, the value of Curtis et al, however, gave only a poor fit.

Table 4.1 Comparison of Microtox Data Obtained in this work with Published Data

Chemical	log ($1/EC_{50}$) from this work	Published EC_{50}	Reference
Acetone	-2.56	-2.56	DeZwart & Sloof, 1983
		-2.58	Surowitz et al, 1987
		-2.57	Curtis et al, 1982
		-2.50	McFeters et al, 1985
		-2.57	Hermens et al, 1985
Toluene	0.61	0.27	Chang et al, 1981
		0.33	Surowitz et al, 1987
		-2.56	McFeters et al, 1985
		0.28	Samak & Noiseux, 1981
1,2,4-trichlorobenzene	2.04	1.89	Ribo & Kaiser, 1983
2-chloronitrobenzene	1.58	1.59	Kaiser & Ribo, 1985
3-chloronitrobenzene	1.17	1.02	Kaiser & Ribo, 1985
4-chloronitrobenzene	0.84	0.88	Kaiser & Ribo, 1985
Butanal	-0.57	0.64	Curtis et al, 1982

4.2.1.1. Cluster Analysis on Variables

Cluster analysis has succeeded in reducing a highly correlated, ill-structured data set of 49 variables into a more ordered form, containing only 14 variables for the Microtox toxicities, and 10 variables for the fathead minnow. The reduction in collinearity is shown in the correlation matrix for the Microtox decorrelated data in Table 4.2. The highest correlation coefficient is now 0.861 (between the Kappa values), as opposed to 0.99 before data reduction (the data matrix of all variables is not shown due to its size).

The level of 90% used to produce the clusters was arrived at in an arbitrary manner. In this case it seems to have been successful, producing clusters expressing significantly different features of the molecule. Also the resulting number of variables fully satisfies the criterion that for stepwise regression analysis there should be no more variables than half the number of cases (Wold and Dunn, 1983). Whether the 90% level should be maintained in future studies is, however, debatable. Livingstone and Rahr (1989) describe an automated system (CORCHOP) that will allow a maximum correlation coefficient of only 0.7 between variables. Without question, though, in this study the method has succeeded in removing a large amount of redundant information.

There are considerable problems in obtaining the best 'representative' variable for each cluster. In this study the choice was made on experience, biased by such factors as ease of calculation, understanding, and which variable was most likely to describe the effect (e.g. size, electronic effect) of the cluster most accurately. For instance, first order valence corrected path molecular connectivity was chosen to represent the large steric cluster, because of its general applicability as a steric parameter. It also models the toxicity of alcohols to the fathead minnow very successfully (see section 3.2.2.14). Table 4.3 shows the relative success of using the other parameters from the steric cluster in the prediction (by regression analysis) of the toxicity of alcohols. The chosen variable is shown to be the most highly correlated to toxicity (r^2(adj)=0.825), thus justifying the selection procedure. However, the majority

Table 4.2 Correlation Matrix for 'decorrelated' Variables in the Microtox Analysis

	CS3	PV1	K2	K3	ClogP	P(S-V)0	C(S-V)3	ClosApp	HOMO	LUMO	MDipole	MHOMO
PV1	0.193											
K2	-0.204	0.813										
K3	-0.048	0.639	0.861									
ClogP	0.192	0.775	0.469	0.238								
P(S-V)0	0.230	0.247	0.284	0.106	0.188							
C(S-V)3	0.585	0.285	0.075	0.059	0.212	0.623						
ClosApp	0.182	0.848	0.793	0.722	0.566	0.243	0.330					
HOMO	0.434	0.509	0.133	0.066	0.546	0.233	0.446	0.401				
LUMO	-0.315	-0.160	0.168	0.264	-0.298	-0.006	-0.068	0.107	-0.403			
MDipole	0.349	0.242	0.160	0.092	0.233	0.497	0.547	0.253	0.124	-0.278		
MHOMO	0.250	0.248	-0.088	-0.222	0.496	0.217	0.357	0.121	0.790	-0.222	0.007	
MLUMO	-0.424	-0.203	0.097	0.224	-0.417	-0.405	-0.334	-0.007	-0.502	0.772	-0.470	-0.314

The abbreviations are those explained in Table 2.5.

Table 4.3 Summary of Statistics for the Relationship of Alcohol Toxicity to the Fathead Minnow, for all the Variables of the Largest Steric Cluster

Variable	Intercept	Slope	s	F	r^2(adj)
MWt	-3.26	0.0258	0.922	36.6	0.652
PS0	-3.64	0.599	0.887	41.6	0.681
PS1	-3.13	0.845	0.842	47.4	0.710
PS2	-2.30	0.841	1.215	13.4	0.395
PS3	-2.16	1.46	0.927	36.0	0.648
PV0	-3.59	0.650	0.730	69.1	0.782
PV1	-3.06	0.933	0.653	90.7	0.825
PV2	-2.48	1.08	0.995	28.8	0.594
PV3	-1.94	1.53	0.860	44.7	0.697
Ka0	-2.25	0.330	0.835	48.6	0.715
K1	-3.30	0.440	0.808	53.0	0.732
K2	-2.20	0.394	0.750	64.4	0.769
K3	-2.23	0.368	0.861	44.6	0.697
Ka1	-3.45	0.461	0.734	68.0	0.780
Ka2	-2.30	0.411	0.677	83.2	0.812
Ka3	-2.28	0.374	0.837	48.2	0.713
P(S+V)0	-3.67	0.317	0.794	55.6	0.742
P(S+V)1	-3.15	0.452	0.731	68.8	0.781
P(S+V)2	-2.50	0.501	1.09	21.3	0.516
P(S+V)3	-2.19	0.808	0.810	52.7	0.731
CMR	-3.34	0.982	0.741	66.5	0.775
ASA	-3.21	0.0114	0.793	55.8	0.743
VWvol	-3.39	0.0279	0.764	61.4	0.761
AltMW	-3.39	0.0205	0.764	61.4	0.761
CollDia	-9.16	1.52	0.770	60.3	0.757
ClosApp	-4.05	0.836	0.885	41.2	0.679
Area	-3.75	0.0248	0.741	66.4	0.775
Volume	-3.55	0.0348	0.721	71.3	0.787
Energy	-3.00	-0.000053	1.082	21.6	0.520
MEnergy	-4.06	-0.000119	0.841	47.6	0.710
Polariz	-2.90	0.0787	0.897	39.6	0.670
MPolariz	-2.67	0.0584	0.966	31.7	0.618

The abbreviations for the variables are taken from Table 2.5. There are 20 observations for each relationship (i.e. n=20)

of the coefficients of determination lie between 0.6 and 0.8, and the overall range is from 0.825 to 0.395 (for second order path molcular connectivity). Clearly, therefore, the choice of best representative variable within a cluster is crucial. It is also reasonable to suggest that the other QSARs obtained might have been improved with the inclusion of different parameters into the original stepwise regression analysis. Livingstone and Rahr (1989) recognise the possibility of removing the 'wrong' parameter by cluster methods, but conclude that it is an acceptable risk with a large number of starting variables. Another check would be to record the correlation of the variables with biological activity at each stage.

4.2.1.2. Description of Clusters

One of the outstanding features of the analysis is the size of the first steric cluster containing the majority of the molecular connectivities. Their high collinearity is to be expected since they are all calculated in a similar fashion. This does mean, of course, that caution should be observed if more than one molecular connectivity term is accepted into a multiple regression eqaution, or if established QSARs employing molecular connectivities are to be utilised. It is also interesting to note that molecular connectivities are highly correlated with both molecular area and volume terms calculated from molecular modelling. Whole molecule polarisability is also included in the steric cluster, thus this measure of polarisability can be assumed to model a steric, and not electronic, effect.

The remaining clusters formed are much smaller and describe other components of molecular structure. Initially considering the variables associated with the fathead minnow data, the second largest cluster comprises the differences between simple and valence path molecular connectivities. The connectivities maybe related to electronic effects, or merely a separate steric feature (a fuller discussion of this group of parameters is given in section 4.2.4.1). Other molecular connectivities found to cluster apart from the purely bulk parameters

are the third order cluster values. These are thought to describe the 'branching' of the non-hydrogen skeleton of the molecule. There are five clusters containing calculated electronic parameters. The dipole moments of the molecule, calculated from the MNDO and CNDO methods have clustered closely together. However, the HOMO and LUMO energies from the different methods (representing the electron donating and accepting capabilities of the molecule) are clustered apart. This maybe due to the differences in the accuracies of both methods - MNDO calculations are presumed to be more accurate. The remaining cluster contains ClogP, the descriptor of the hydrophobic element of the molecule.

Cluster analysis of the variables associated with the Microtox data gave a very similar pattern in the grouping of the descriptors. There are however three extra clusters formed, containing a total of five variables. (These variables were found in the large steric cluster in the cluster analysis of the variables associated with the fathead minnow data.) Four Kappa indices have clustered apart from the main steric cluster. This is probably a consequence of their describing more distinct elements of the shape, or symmetry of the molecule, as opposed to being pure bulk terms. The other descriptor clustered apart is the closest approach which may be quantifying features of a molecule's surface area, or overlap, that are less related to its bulk. The exact reasons for the differences in the results of the cluster analyses between the two data sets are, however, difficult to ascertain, but will be as a result of the intrinsic chemical variability between the chemical structures present in each data set.

4.2.2. Analysis of QSAR Techniques

4.2.2.1. Prediction of Microtox Toxicity

The whole set of compounds is moderately modelled by hydrophobicity alone (r^2(adj)=0.581; eqn 3.1). This relationship is slightly improved by the addition of two more significant parameters, namely CNDO calculated HOMO energy and second order Kappa value. Because of the low correlation and the decrease in stability as additional parameters are included (expressed as a fall in the F-statistic) of equations 3.1-3.3, caution should be employed if they are to be used for predictive purposes.

More encouraging relationships are encountered for individual classes. Good correlations are observed for the esters and the alcohols, equations 3.7 and 3.16 both being based on hydrophobicity and a steric term. The prominence of the hydrophobic descriptor is to be expected as both classes are thought to act as simple narcotic agents, and as such their toxicity should be well modelled by hydrophobicity (Konemann, 1981). Also work such as that of Leegwater (1989) (see Appendix 2) and Protic and Sabljic (1989) has proven the utility of steric terms, such as molecular connectivities, to model narcotic toxicity.

The equations for the toxicities of each of the chemical classes with ClogP alone are shown in Table 4.4 in order to show the relative toxic effect compared with hydrophobicity. Compounds acting by a purely narcotic mechanism should have a similar intercept (approximately minus two) and slope (approximately one) to that reported by Hermens et al (1985) for 22 organic chemicals (including alcohols, chlorinated alkanes, and chlorobenzene derivatives) thought to be acting as simple narcotics in the Microtox test:

$$\log(1/EC_{50})_M = 0.995 \log P - 2.14 \qquad (4.2)$$

n=22 s=0.53 r=0.952 F not given

N.B. This equation was obtained using toxicity data taken after 15 minutes. Such data have been found to be easily comparable with 5 minute data (see section 4.1).

Table 4.4 Summary of the Statistics of the Relationships of Microtox and Fathead Minnow Toxicity with ClogP for each Chemical Class Considered

i) Microtox Toxicities

Class	Intercept	Slope	n	s	F	r^2(adj)	
Ketone	-1.02	0.738	13	0.999	9.1	0.403	
Ester	-1.96	1.113	11	0.470	106	0.913	*
Nitrile	-0.935	0.894	7	1.103	8.2	0.545	
Aldehyde	0.733	0.079	8	0.904	0.1	0.000	
Amine	0.508	0.152	8	0.398	2.8	0.208	*
Benzene subs	0.244	0.409	10	0.488	2.6	0.150	
Alcohol	-2.05	1.14	14	0.766	30.7	0.696	

ii) Fathead Minnow Toxicities

Ketone	-1.20	0.682	13	0.632	19.4	0.606	
Ester	-0.012	0.555	16	0.518	31.8	0.672	*
Nitrile	0.133	0.410	8	1.513	2.3	0.153	
Aldehyde	1.74	-0.088	8	1.112	0.1	0.000	
Amine	-0.560	0.545	8	0.361	44.1	0.860	*
Alcohol	-1.44	0.789	20	0.735	67.8	0.779	

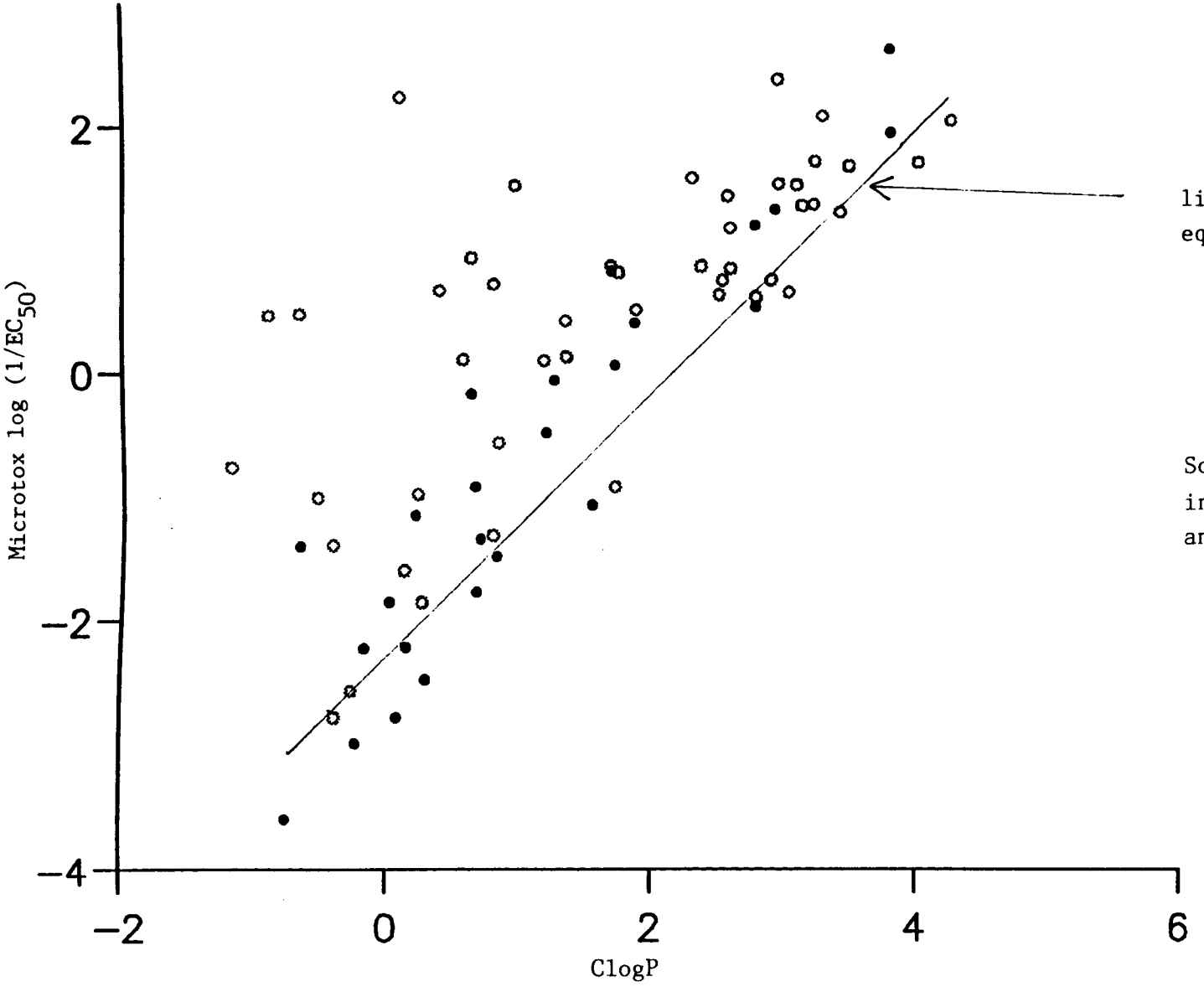
* Indicates that the relationship with ClogP was the most significant one-variable equation obtained

---

It is obvious that the relationships with ClogP for the esters and alcohols reveal equations very similar to eqn 4.2, and thus these compounds can be considered to act by a similar mechanism. Figure 4.2 shows all the compounds in the data set plotted against ClogP, the fitted regression line of eqn 4.2 giving a good approximation of baseline toxicity, modelling well the toxicity of the alcohols and esters. Fig 4.2 also suggests that the majority of the chemicals in the other classes are acting by mechanisms other than simple narcosis, as they occur above the fitted line of eqn 4.2.

Ketones, which are thought to be narcotics, are also moderately well related to a steric term. The relationship with hydrophobicity (see Table 4.4) gives only a poor correlation however, and the intercept is one log unit above that expected for a narcotic relationship. Surprisingly, due to it's seemingly unreactive nature, the most significant outlier is 3,3-dimethyl-2-butanone, and it can

Figure 4.2 Graph of Microtox toxicity for all compounds used in the QSAR analysis against ClogP



therefore be concluded that some of the ketones considered in this study act by a more specific mechanism, at least in the Microtox test.

A moderately good correlation was found for the toxicity of the nitriles (eqn 3.9). Again a steric parameter was found to be the best descriptor (r^2(adj)=0.749), the relationship with ClogP (see Table 4.4) being only poor. As would be expected for reactive compounds such as the nitriles (nitriles are known, for instance, to metabolise to cyanide (Tanii and Hashimoto, 1984)), the poor relationship with hydrophopbicity suggests that other toxicity mechanisms are also operating.

The amines and the aldehydes are very poorly modelled; indeed the F-statistic in neither eqn 3.10 nor eqn 3.12 is significant at the 95% level (Pearson and Hartley, 1972). It is difficult to draw conclusions about the mode of action of the amines from the relationship with hydrophobicity. It is likely, for instance, that aliphatic amines will act as simple non-polar narcotics, whereas anilines may act by polar narcosis.

When the toxicity- log P relationship is studied for the aldehydes, one obvious outlier becomes apparent, namely acrolein. This is known to have substantially increased toxic effects because of its virtually unique property to act a Michael acceptor (Lipnick et al, 1987). When acrolein is removed from the data set a much improved relationship is seen with ClogP:

$$\log(1/EC_{50})_M = 0.725(0.13)\ \text{ClogP} - 0.831 \qquad (4.3)$$

n=7 s=0.287 r^2(adj)=0.826 F=29.5

This is consistent with the finding of Deneer et al (1988) that although aldehydes may act by a mechanism other than narcosis (reacting with nucleophilic entities by addition reactions), there is still a strong correlation of their toxicity with hydrophobicity. The fact that the aldehydes are not acting by narcosis is confirmed by the intercept of the regression line in eqn 4.3 being considerably above that of eqn 4.2.

N.B. Equations 3.11 for the aldehydes and 3.13 for the amines must be discarded for predictive purposes as they contravene the Topliss and Costello rule (1972).

This states that more than 1 variable to each 5 observations in a regression equation will create a significant risk of chance correlation, thus equations with less than about 10 observations can strictly justify only one descriptor. The last chemical class considered, the substituted benzenes, is very poorly modelled. The relationship with ClogP is not significant, and this may be a result of the substituted benzenes being a very heterogeneous group of compounds (including, for instance, chloro, and nitro substituents, as well as toluene). As such many different toxicity mechanisms will occur within the class. The chlorobenzenes for instance, may act purely by narcosis, in contrast to the nitro-substituted aromatics which may be metabolised to more reactive intermediates. Different toxicity mechanisms will always give rise to problems in the modelling of the data by QSAR mechanisms, and a poor relationship with hydrophobicity.

4.2.2.2. Prediction of Fathead Minnow Toxicity

For all the classes of chemicals considered together (eqns 3.18-3.23) there is only a moderate relationship, mostly influenced by ClogP, and not suitable for predictive purposes. Of interest, though, is the fact that best subsets regression analysis has obtained the best equations containing 3 or 4 variables. There is an increase in the variance explained of 0.3% and 1.0% respectively, a decrease in the standard error and a slight increase in the F-statistic for eqns 3.22 and 3.23 from best subsets regression analysis. The statistical gains may therefore be considered negligible in this example, but it does perhaps suggest that best subsets regression analysis should be used instead of stepwise regression whenever possible. Experience with this technique has shown a disadvantage of best subsets regression to be an increase in the amount of computer time needed to achieve a result.

Within individual chemical classes, good relationships were obtained for the ketones (although eqn 3.26 must be discarded according to the Topliss and Costello rule), amines and alcohols. Again the utility of the molecular

connectivities as QSAR parameters is demonstrated, with good correlations with ketone and alcohol toxicities. Table 4.4 reports the relationships of the toxicities of each of the chemical classes with ClogP alone. The alcohols (and to some extent the ketones) are the only classes apparently acting solely by narcosis, having an intercept of approximately -1.5, and a slope of approximately 0.8 in the relationship with ClogP. This is similar to that reported by Cronin and Dearden (1990) for the toxicity to the fathead minnow of 21 chlorinated aliphatic and aromatic compounds thought to act as simple narcotics:

$$\log(1/LC_{50})_{FM} = 0.829\ ClogP - 1.48 \qquad (4.4)$$

n=21 s=0.207 r^2(adj)=0.944 F not given

Fig 4.3 shows the toxicities of all the chemical classes plotted against ClogP. The fitted regression line for eqn 4.4 is a good model of baseline narcosis and a good model of alcohol toxicity; exceptions to this are 2,3-dibromopropanol and 2-chloroethanol. The reasons for these to appear as outliers are unclear, Purdy (1987) reports that alcohols can be enzymatically oxidised to aldehydes, but there is no explanation for oxidation not to occur for all alcohols, or why it has occurred only for the two mentioned (especially when other chlorinated alcohols are observed to act by a purely narcotic mechanism).

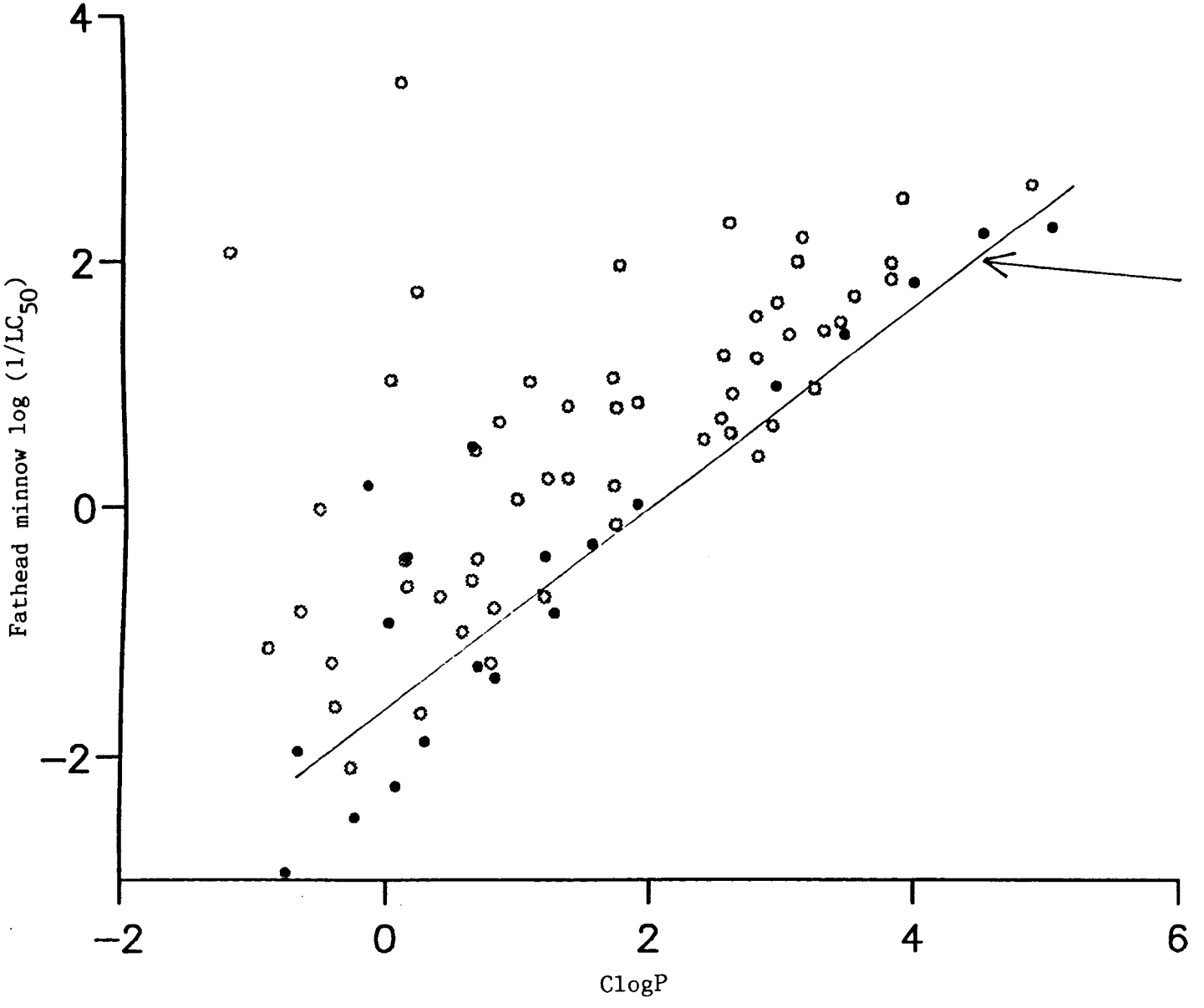
The good relationship between amine toxicity and ClogP (eqn 3.32) indicates that these chemicals are not acting by narcosis, but may well be acting by polar narcosis. Equation 3.32, for instance, resembles the relationship (eqn 4.5) established by Veith and Broderius (1987) for polar narcosis in the fathead minnow (i.e. a decrease in the slope, and in increase in intercept as compared to eqn 4.4):

$$\log(1/LC_{50})_{FM} = 0.65(0.07)\ \log P - 0.71 \qquad (4.5)$$

n=39 s not given r^2(adj)=0.90 F not given

The esters are reasonably well modelled by a two parameter equation (eqn 3.22), and the poor relationship with hydrophobicity (eqn 3.27) indicates that more specific toxicity mechanisms may be operating. Veith et al (1985) also found esters to be more toxic than predicted by narcosis in the fathead minnow,

Figure 4.3 Graph of fathead minnow toxicity for all compounds used in the QSAR analysis against ClogP



suggesting that the more polar esters may act as polar narcotics.

The QSARs for the nitriles and aldehydes were very poor, particularly with respect to ClogP. When the aldehydes were re-analysed with acrolein removed, only a small improvement was found:

$$\log(1/LC_{50})_{FM} = 0.658(0.25)\ ClogP - 0.0694 \qquad (4.6)$$

n=7 s=0.535 r^2(adj)=0.499 F=7.0

This does not compare well with the relationship with hydrophobicity found for Microtox and may be a result of the longer test time in the fish assay (96 hours as compared with 5 minutes) allowing more specific toxic reactions to occur. As a consequence of this complex mode of toxic action, the toxicity of these compounds in the fathead minnow may be more difficult to predict. The greater relative effects of more reactive compounds are also found in the inter-species relationship between fathead minnow and Microtox (see section 4.3.1.1), where increased toxicity is observed especially with the fathead minnow. This phenomenon seems more likely to occur with reactive chemical classes, such as the aldehydes and nitriles, than those acting by non-polar, or polar, narcosis.

4.2.2.3. Comparison of QSARs for Microtox and Fathead Minnow

Hydrophobicity is the most important chemical descriptor when all the chemical classes are considered together for both species. There was a better correlation with the Microtox, than with the fathead minnow toxicity data. In both cases the ClogP equations were also improved by the inclusion of calculated electronic parameters. The end results were still, however, disappointing for the two species, indicating that the compounds in both data sets were not acting by a narcotic mechanism alone. Indeed both eqns 3.1 and 3.18 are significantly different from the baseline toxicity equations (eqns 4.2 and 4.4) for each species respectively.

The results show it is definitely more profitable to study a homogenous data set as opposed to a chemically heterogeneous one. However, it was only for the alcohols that good correlations were obtained for both species. Generally, good

relationships are observed for chemical classes that are believed to be unreactive and thus to act by a simple narcotic mechanism; these were well modelled by hydrophobicity (e.g. eqn 3.6), or by steric parameters (e.g. eqns 3.16 and 3.32). Some classes expected to act as simple narcotics were, however, modelled only poorly in one or both of the species. Both for the ketones and amines considered in the data, good correlations were found for the fathead minnow, yet not for the Microtox toxicities. Of these, it was postulated that the amines were acting by polar narcosis (Schultz et al, 1990b), a better relationship occurring with the fathead minnow data because the longer test time has allowed the full toxic effect to occur more completely (i.e. a state state has not been achieved in the Microtox test).

More reactive compounds such as the aldehydes and the nitriles were less well modelled by any of the parameters used in this work, even though several parameters could be considered to model reactivity.

An obvious feature of the QSARs is the importance of hydrophobicity (ClogP) and steric terms (predominantly first order valence-corrected molecular connectivity). The good fit of hydrophobicity to much of the toxicity data confirms the supposition (see section 1.6.7) that the major factor in determining a compounds toxicity is its transport to the active site. Thus for compounds with a similar mode of toxic action, it is their partitioning behaviour (which governs transport) which is the 'rate limiting step' of toxicity. Molecular connectivities model the toxicity data well in several cases, however since they do not parameterise hydrophobicity, they cannot strictly be related to a molecule's transport, but some other steric effect that is important in toxicity. The calculated electronic terms are found in the equations for the more reactive chemical classes. This suggests that they are modelling some attribute of the chemical reactivity e.g. these electronic effects may be important in the metabolism of the chemicals.

In this study the modelling of the toxicity of the aldehydes is an example of the problems that can be associated with regression analysis for predictive purposes.

Initially no satisfactory relationship was observed, but the removal of one obvious outlier, not identified by stepwise regression analysis (except as an unusual observation with a large residual), gave a remarkable improvement in the result. This shows the value of a graphical, or other visual, representation of the results.

N.B. Unfortunately there were too few data to perform a comparative QSAR study on the Daphnia magna (twelve values) and Tetrahymena pyriformis (three values) toxicities.

4.2.3.1. Principal Component Analysis

Principal component analysis has successfully reduced the data set of 49 highly collinear variables, into five significant new variables. These principal components are fully orthogonal and account for most (92.7% for the Microtox data and 93.7% for the fathead minnow) of the total variance of the original data. Thus the bulk of the information content of the multi-dimensional data space has been explained by the five principal components.

Predictably, the structure of the results of the principal component analysis is similar for both the fathead minnow and Microtox data sets. The principal components are rather ill-defined and it is difficult to establish if any 'chemical meaning' can be attached to them. The variables with the highest loadings are summarised in Tables 3.5 and 3.7. The first principal component appears to represent size or bulk. All the molecular connectivities (except for the cluster values) and the COSMIC steric terms have a very similar loading (0.160 to 0.178), so the PC cannot be defined more precisely. The second principal component accentuates the difference between HOMO and LUMO (i.e. the electron donating and accepting capability of the molecule). Not only do the difference in HOMO and LUMO energies (DiffH-L and MDiffH-L) variables have high loading, but so do HOMO and LUMO energies themselves, with the loading for LUMO being positive, and that for HOMO being negative. Also, for the Microtox data only, there is an emphasis of the second and third order Kappa values which describe the shape or symmetry of the molecule. Prominent in the third principal component are the cluster and difference in simple and valence molecular connectivities, suggesting another steric element. In addition, for the fathead minnow there are high loadings for LUMO and HOMO-LUMO difference. The fourth principal component is dominated by cluster molecular connectivities, which account for branching of the non-hydrogen skeleton of the molecule. This PC also contains (with lower loadings) dipole moments and HOMOs. The fifth principal component emphasises electronic characteristics, especially the difference between the dipole moment (with a high negative loading) and HOMO and LUMO

energies (with high positive loading).

The results show that within the data, some variables, such as cluster molecular connectivities, HOMO, LUMO, and dipole moment have a more significant role to play in analysing the information content than have other variables. However, the features that the analysis extracts from the data may not be relevant to a QSAR study, because they may not describe the pertinent molecular features which are important in explaining toxicity. This can be considered to be a general drawback of such 'unsupervised learning' statistical methods. An example is that although ClogP was found in each of principal components (with the highest relative score in PC1) there is an absence of a truly hydrophobic principal component, which is, of course, a very important factor in any QSAR study.

4.2.3.2. Principal Components in the QSAR Analysis

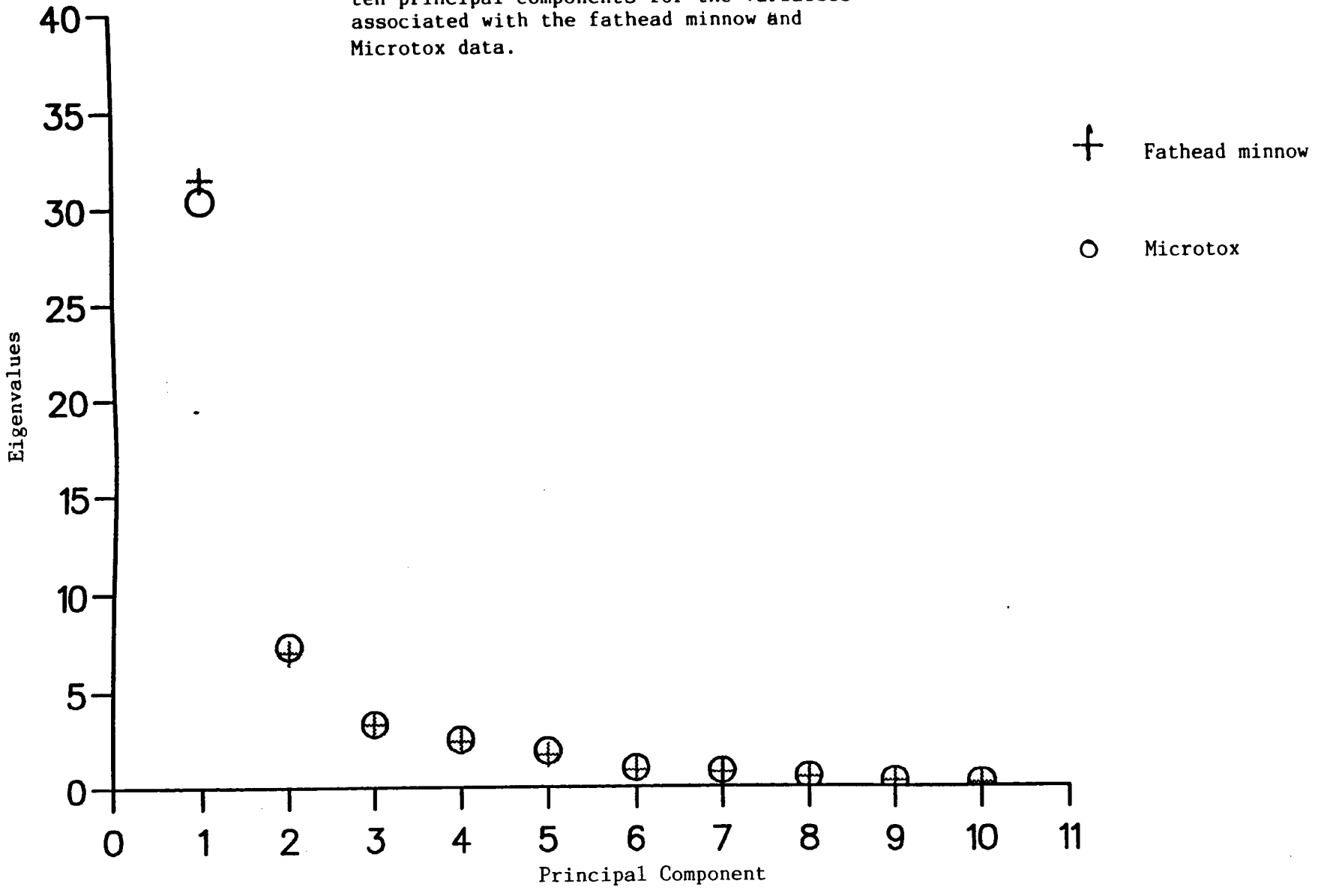
The results from the QSAR analysis using the principal components are disappointing. Of the five principal components only three are found to be significantly correlated with the toxicities. For the Microtox study, toxicity in the final equation (3.40) is positively correlated to the steric component (PC1), yet negatively correlated to some measure of the branching (PC4) and an electronic effect (PC2). The fathead minnow toxicity data (eqn 3.43) are also positively correlated to PC1, but again negatively correlated to branching (PC4) and an electronic effect (PC5). The appearance of negative correlations is somewhat of a surprise, and perhaps concern in the QSARs validity. It would perhaps be expected that features such as the branching of the molecular structure (i.e. basically a steric descriptor) are positively correlated with toxicity due to its effect on the transport of a molecule, thus an increase in branching will lead to an increase in toxicity. Conversely, if steric hindrance is a factor affecting, for instance, a specific toxicity mechanism, the size or bulk, of the molecule may become a limiting factor in its toxicity. Fisher et al (1987), for instance, observed that molecular weight was inversely related to the nasal absorption of water-soluble compounds in the rat, and it was concluded that

whilst partitioning controls the major part of drug permeation across membranes, the sieving of water-soluble molecules via aqueous shunts or possibly other mechanisms should not be neglected. This may be one of the shortcomings of using principal components when not only their meaning is very hard to determine, but also that of the original variables. More proof of the lack of stability in the QSARs is the reduction in the F-value of the equations when more variables are added.

4.2.3.3. Significance of the Principal Components

A simple test for the significance of the principal components is a scree plot of the eigenvalues against principal component number. Fig 4.4 shows that a plateau is reached after the fifth principal component, confirming that the value of one for testing the significance of the eigenvalue is applicable.

Figure 4.4 Scree plot of the eigenvalues of the first ten principal components for the variables associated with the fathead minnow and Microtox data.



4.2.4.1. Factor Analysis

In common with principal component analysis there is close agreement between the two sets of factors for both data sets. Factor analysis has succeeded in eliminating collinearity by the calculation of 5 new factors that account for most of the variance (92.7% for the Microtox data, and 93.6% for the fathead minnow).

The first factor is a general descriptor of size and bulk. Many parameters are 'highlighted' (i.e. have a high loading), including the classic steric descriptors such as measures of area and volume, as well as the molecular connectivities. The second factor is also concerned with molecular size, but a much more specific attribute of the molecule, as described by the difference between simple and valence corrected molecular connectivities. The exact meaning of this factor is very difficult to determine. Kier and Hall (1986), for instance, believe that it is related to electronic effects, however Dearden et al (1988) found no correlation with electronic parameters, and only a poor correlation with steric parameters. The third factor is a purely electronic feature of the molecule, emphasising the LUMO energy, or the ability of the molecule to accept electrons. The fourth factor is dominated by cluster molecular connectivities which are known to encode the branching within the molecule. Electron donation is represented by the fifth factor which is highly reliant on the HOMO energy. Also, for the Microtox data, the dipole moment is important. This may, of course, have an effect on the electron donating properties of a molecule.

4.2.4.2. Using the Factors in QSAR Analysis

When the scores for the chemicals calculated from the individual factors were considered in the QSAR analysis, only disappointing relationships are obtained. For both toxicities significant equations were obtained with up to four factors being used. The Microtox data were most correlated to the first, largely steric, factor, followed by a positive correlation with the electron acceptance descriptor factor 3. The addition of the second and fifth factors improved the

equations slightly, but the overwhelming feature of the addition of more variables into these equations was a reduction in their stability, indicated by significant decreases in the F-value. Similarly the fathead minnow toxicities are most correlated to factors 1 and 3, but in this case there is a negative correlation between toxicity and factor 3. The reason for this negative correlation is not apparent. Toxicity is also negatively correlated to branching (factor 4), which may be due to the effect of size inhibiting transport across a membrane (see section 4.2.3.2). Yet again there seems to be little overall stability to these equations.

4.2.4.3. Principal Component Analysis vs Factor Analysis

The results from both analyses are, as expected, closely related. Five significant principal components and factors were obtained, accounting for approximately 93% of the variance in the original data. Broadly speaking, the information of these new variables describes two steric properties, the intramolecular branching, and electron acceptance and donation capabilities of the molecules. Principal component analysis gave more significance (i.e. a higher eigenvalue) to electron acceptance than did factor analysis. Within the structure of newly calculated variables the factor analysis has the distinct advantage that through the rotation of the factor axes, considerable clarification of the chemical meaning of each factor has been achieved. The difficulty in defining the principal components, as well as a simple comparison of the unrotated and rotated factor loadings confirms, this conclusion. On this subject, Harris (p223, 1975) reports that because of the intrinsic loss of 'contact' between the original and new variables, principal component analysis is greatly preferable to factor analysis. However, it must be noted in this study that factor analysis has given much more 'discrete' meanings to the variables, and thus has aided in their elucidation.

Unfortunately neither set of parameters has performed particularly well in the QSAR analysis. The equations obtained have poor predictive power (a maximum of

61% of the variance explained) and lack stability. In both analyses, the steric component of the data has been the best model, in the absence of a hydrophobic descriptor. However, there are also negative correlations with the intra-molecular branching that may be modelling the effect of steric hindrance in membrane permeability.

4.2.5.1. Canonical Correlation Analysis

Canonical correlation analysis has successfully aided in the quantification and understanding of the relationship between the two sets of variables. By the production of two new sets of two orthogonal parameters, representing toxicological and physico-chemical data space, much of the variation between the data has been explained. Reasonable canonical correlations of 0.837 and 0.712 have been achieved.

There are strong correlations between the two toxicities and CNVRF1, (see Figs 4.5 and 4.6). CNVRF1 can be considered as a general expression of aquatic toxicity (see Section 3.2.5.3), being a function (or the weighted sum) of both biological activities. CNVR2 is, however, expressing factors that may contribute to differences in the relative toxicity of the compounds i.e. a negative standardised coefficient for the Microtox data, and a positive value for the fathead minnow data. The weighted difference in the toxicities will identify compounds with a similar relative toxic effect (i.e. they will have a score for CNVRF2 of approximately 0), yet those relatively more toxic to one species or another will have a score significantly different from 0 (higher scores for those chemicals more toxic to fathead minnow, and lower scores for chemicals more toxic to Microtox). Section 3.3.1 indicates the chemicals that are relatively more toxic to either fathead minnow or Microtox and these are listed in Table 3.17. Figs 4.7 and 4.8 are bivariate plots of the two canonical variables for the first (toxicity) and second (molecular descriptor) data sets respectively plotted against each other. These are thought to represent the toxicological and phyico-chemical 'hyperspace' of the data. In addition the chemicals are scored according

Figure 4.5 Graph of Microtox toxicity against the first canonical variable of the first data set (CNVRF1)

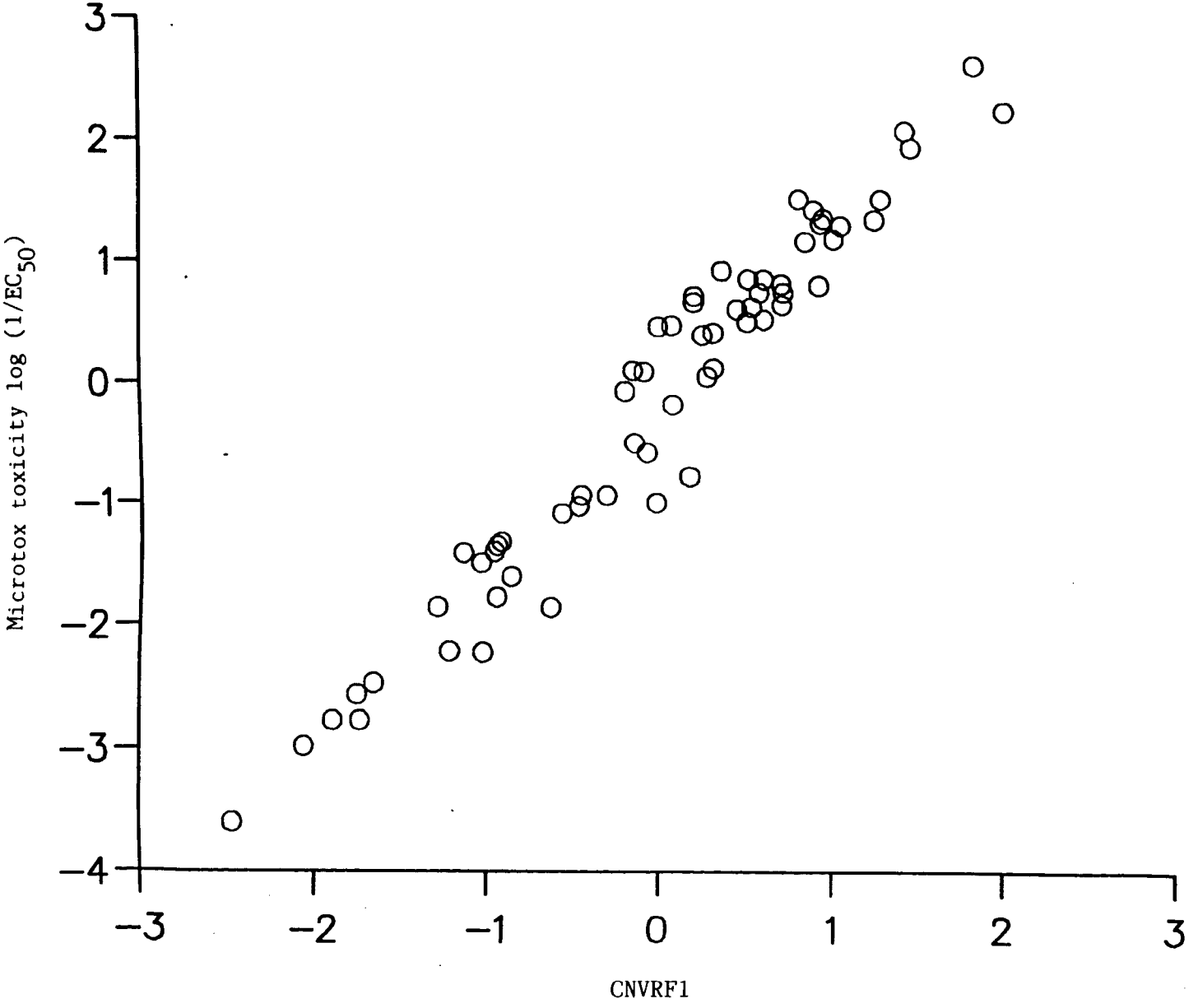


Figure 4.6 Graph of fathead minnow toxicity against the first canonical variable of the first data set (CNVRF1)

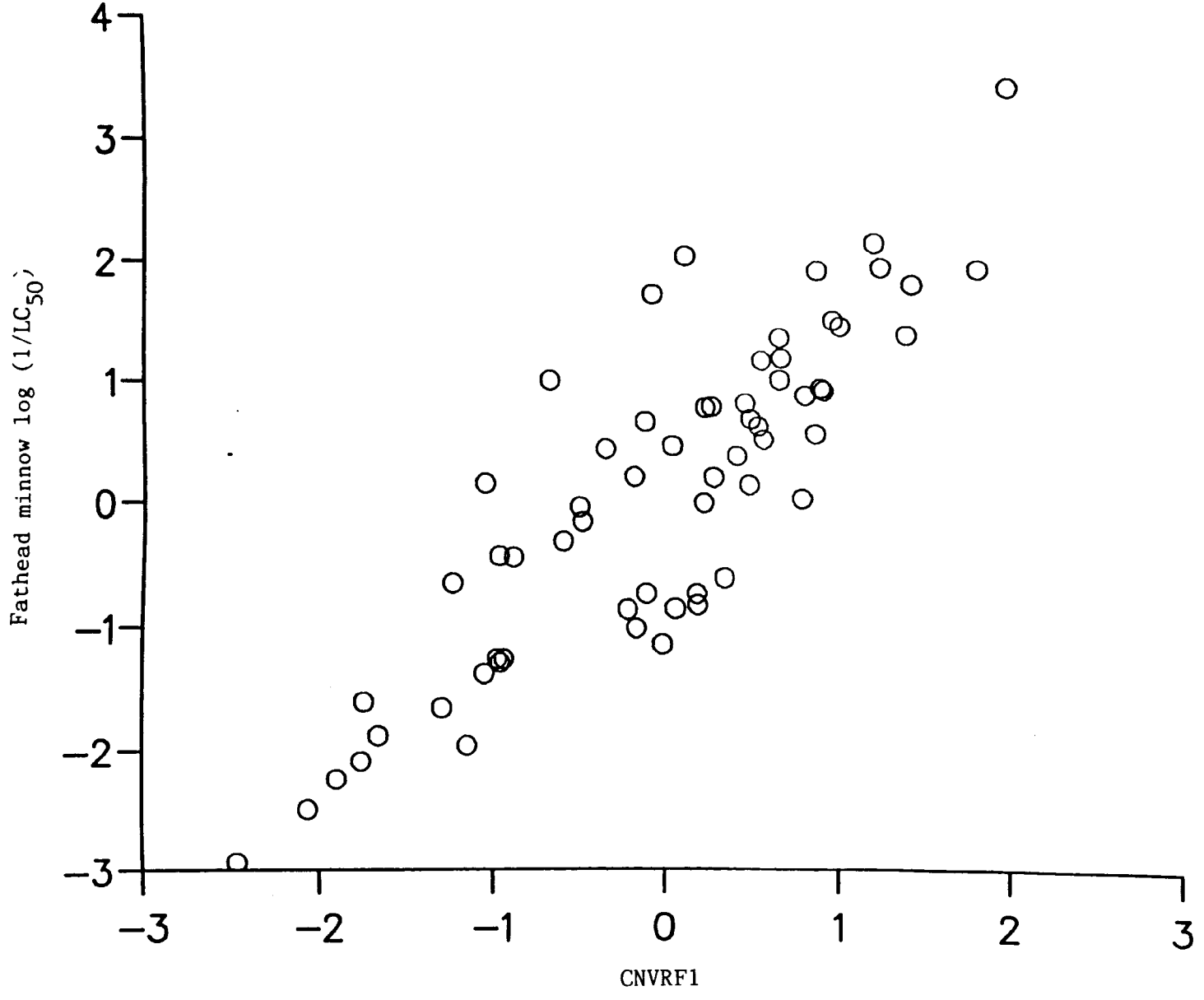


Figure 4.7 Plot of the scores for the first (CNVRF1) and second (CNVRF2) canonical variates of the first data set. (Relative toxicities taken from Section 3.3.1.)

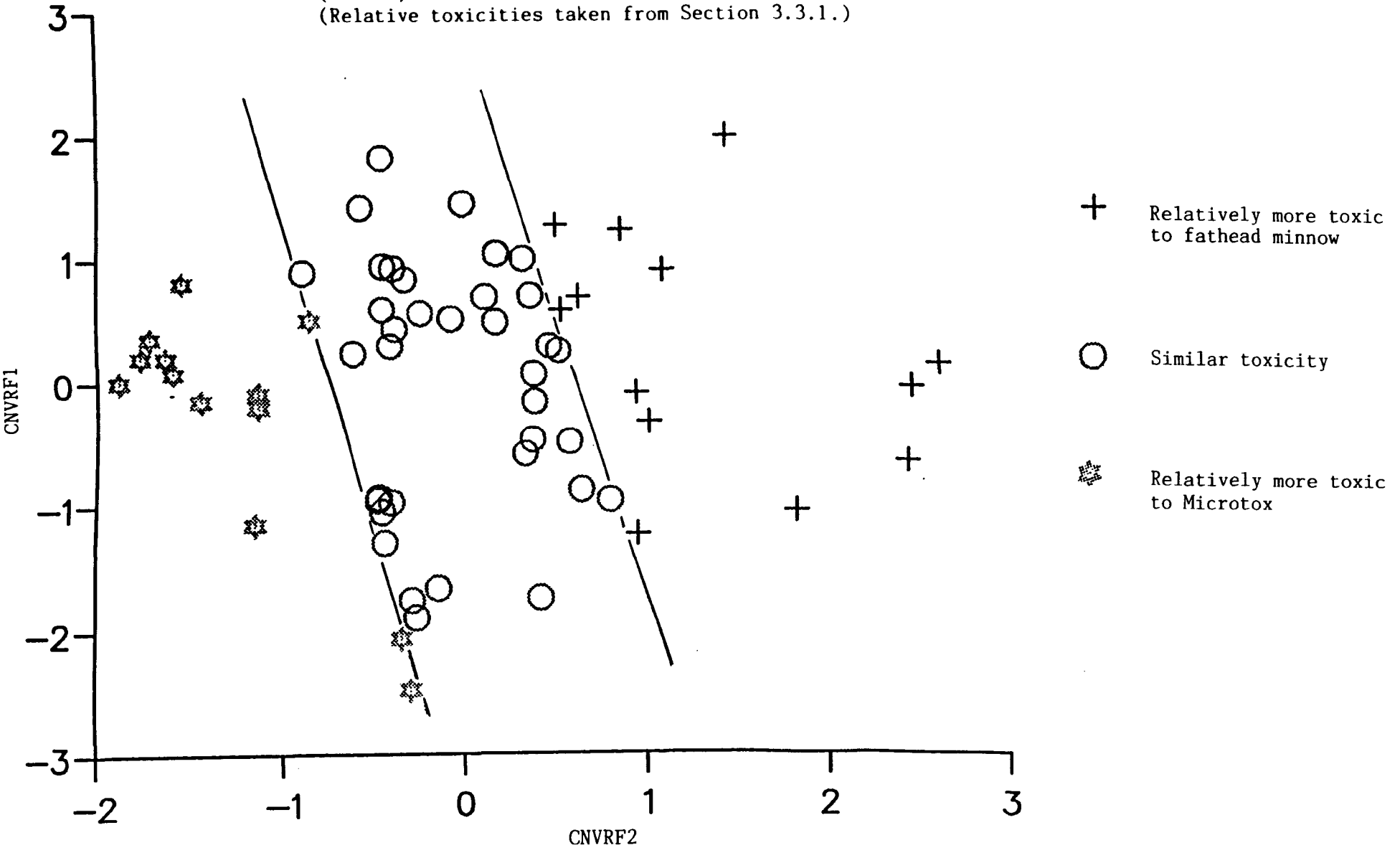
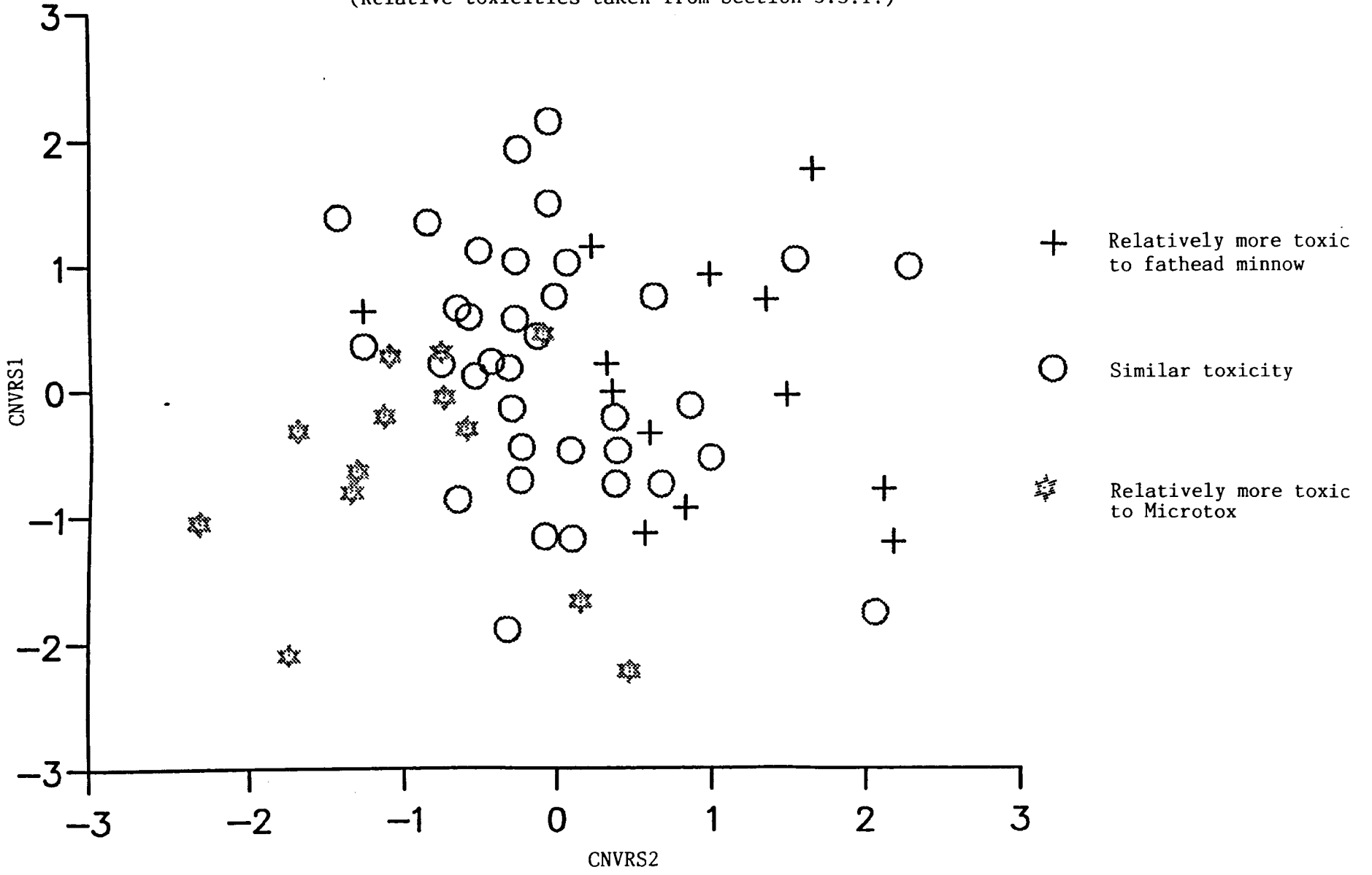


Figure 4.8 Plot of the scores for the first (CNVRS1) and second (CNVRS2) canonical variates of the second data set. (Relative toxicities taken from Section 3.3.1.)



to the relative strength of their toxicity to the two species. Thus in Fig 4.7, there is a clear association between the relative toxic effect and its score for CNVRF2. The boundaries of each area of the graph are not perpendicular to the x-axis of the graph due to the slightly higher standardised coefficient of the fathead minnow toxicity in the calculation of CNVRF2 (1.433) compared to the Microtox (-1.276). Fig 4.7 can be thus considered as a graphical expression of the difference in the relative toxicities. Fig 4.8 displays this effect less clearly, due to the lack of predictive power of the descriptor CNVRS. However, all compounds more toxic to the fathead minnow are on the right of the graph and those more toxic to the Microtox on the left.

4.2.5.2. Physical and Chemical Meaning of the Canonical Variates

Analysis of the descriptors in the canonical variables for the second set gives an indication of the molecular features that may influence the effects accounted for by the first set. Thus analysis of CNVRS1 will show features leading to a general toxicity estimate. The standardised coefficients for all the canonical variables are shown in Section 3.2.5.3. The larger the coefficient the more influence the descriptor has on the resulting canonical variable. CNVRS1 has high coefficients for HOMO energy (0.504) and ClogP (0.410). This combination of an electronic and hydrophobic parameter for predicting aquatic toxicity has of course been reported many times before. Steric descriptors PV1 (0.244) and P(S-V)0 (0.13) are also positively associated with toxicity. Toxicity is, however, negatively correlated with branching, as modelled by the molecular connectivities C(S-V)3 (-0.373), and CS3 (-0.114). This is similar to the phenomenon observed in the principal component and factor analyses, when toxicity was negatively correlated with the principal component and factor encoding molecular branching. Of the remaining electronic descriptors, only the calculated dipole moment is marginally important (0.166), the other electronic descriptors being largely insignificant.

CNVRS2 is attempting to describe the difference in the relative toxicological effects. Thus a high positive value for CNVRS2 (indicating greater relative

toxicity to fathead minnow) results from chemicals with high values for those descriptors with high positive coefficients. These are predominantly ClogP (0.469), P(S-V)0 (0.284), and C(S-V)3 (0.293). A slight contribution from dipole moment (0.164) is also observed. Greater relative toxicity to Microtox is expressed by the electron accepting and donating capabilities of a compound, LUMO (-0.611), HOMO (-0.568), MHOMO (-0.296), MLUMO (-0.256). The effect of CNDO calculated parameters was found to be about twice that of those calculated by the MNDO method. Interestingly, two molecular connectivities are important PV1 (-0.491), and CS3 (-0.376), and these indicate that there is a large steric component in CNVRS2.

One has to be very careful drawing conclusions from this evidence, not least because only 50.7% of the variation of CNVRF2 is actually explained by CNVRS2. However, it may be that relatively hydrophobic molecules with small electronic effects (as measured by HOMO and LUMO) are more toxic to the fathead minnow, whereas large molecules, with greater electronic effects will exhibit a greater toxic response in the Microtox bioassay. A simple test of these theories is analysis of the corresponding descriptors and chemical groups. Table 4.5 shows the mean values of the CNDO and MNDO calculated HOMO and LUMO energies, and ClogP for the data involved in the canonical correlation, considered by chemical class. Chemical classes of interest are the aldehydes (thought to have relatively greater toxicity to the fathead minnow - see section 4.3.1) and the ketones and alcohols (to which the Microtox test species may be relatively more susceptible). The mean of the LUMO energies is lower for the aldehydes (CNDO 1.110; MNDO -0.425) than compared to that for the ketones (CNDO 3.958; MNDO 0.711), and the alcohols (CNDO 4.97; MNDO 3.00). There is, however, a less clear separation of the values for the HOMO energies but these do show that the aldehydes have greater mean values (CNDO -12.718; MNDO -11.906) than the ketones (CNDO -12.607; MNDO -12.516), and the alcohols (CNDO -14.210; MNDO -13.524). Thus the ability of the aldehydes to donate electrons may be an important factor in their increased

Table 4.5 Arithmetic means for ClogP, HOMOs and LUMOs for the data used in the canonical correlation analysis

Chemical Class	ClogP	HOMO	LUMO	MHOMO	MLUMO
Ketone n=13	1.224	-12.607	3.958	-12.516	0.711
Ester n=11	1.757	-13.289	4.723	-13.120	1.302
Nitrile n=7	0.324	-14.514	4.859	-14.465	1.230
Aldehyde n=8	1.841	-12.718	1.110	-11.906	-0.425
Amine/Aniline n=8	1.442	-12.816	5.700	-11.671	2.330
Benzene/Toluene n=2	2.694	-12.960	2.250	-11.767	-0.850
Alcohols n=13	0.638	-14.210	4.970	-13.524	3.004
All Classes n=62	1.249	-13.332	4.210	-12.843	1.367

toxicity to the fathead minnow, relative to the Microtox test, (according to the canonical correlation analysis). Also the aldehydes in this study have a higher mean ClogP than the ketones or alcohols as described by the analysis.

These results may also be explained by the bioaccumulation potential of a chemical. The bioaccumulation (or bioconcentration) of chemicals is dependent on factors such as time and directly related to hydrophobicity (see section 1.6.6). Thus bioaccumulation will be much greater in the fathead minnow test due to the longer testing time. In addition the mean ClogP for the aldehydes is higher than, for instance, the alcohols, so it could be reasoned that overall the aldehydes will accumulate more than the alcohols. Both these facts mean that if toxicity is dependent on the accumulation of the chemical, the toxic effect of chemical classes such as the aldehydes will be more pronounced in the fathead minnow.

4.2.5.3. Prediction of Toxicity using Canonical Variates

The results of the canonical correlation analysis can also be used to predict, or extrapolate, toxicity. For the first variate the following relationship exists:

$$CNVRF1 = 0.837(0.071)\ CNVRS1 - 0.0001 \qquad (4.7)$$

$n=62 \qquad s=0.552 \qquad r^2(adj)=0.696 \qquad F=140$

As already explained, the descriptors needed to calculate CNVRS1 are very easily obtained. Thus the equation can be solved for CNVRF1 if one of the toxicities is known. Although only 70% of the variance is explained by equation 4.7, this is approximately 6% more than was described by a comparative conventional QSAR obtained from multiple linear regression analysis (e.g. eqns 3.3 and 3.23). Also due to the nature of this technique, i.e. the inclusion of contributions from many different parameters, the results are likely to be more stable than those procured from regression analysis.

4.2.6. Validation of the Various QSAR Methods Using a Testing Set

4.2.6.1. Prediction of Microtox Toxicities

The results for the predicted values of Microtox toxicity are listed in Table 3.15. There are interesting variations of the results within the methods; the results for the QSAR calculated from all the classes give only reasonable estimates of toxicity for all the chemicals, but none is a truely accurate, or reliable, representation of the toxicity. There is no overall pattern to the discrepancies in observed and calculated toxicity, and the QSAR used to calculate the toxicities can be considered as a stable equation containing the classic QSAR descriptors (hydrophobic, steric, and electronic). The reasons for the disappointing predictive performance of the equation are therefore difficult to elucidate.

When predictions of toxicity are considered from the QSARs for individual classes, good results are found for ethylpropionate and heptanol for which QSAR models with high correlation coefficients were obtained. The toxicities of 4-chlorobenzaldehyde and 2,3-dichloronitrobenzene are also well estimated, despite the QSARs being very weak. Calculated values for chloroacetone, 4-chlorobenzonitrile, and butylamine are very poor, the last being the product of a highly unstable, insignificant QSAR. The results from such relationships must be treated with extreme caution, and assumed to be spurious. It is obvious that the strength, and quality of the relationship will dictate the accuracy of any prediction made from it.

Of the predictions from the regression analysis on principal components only those for chloroacetone and 2,3-dichloronitrobenzene have provided a reasonable estimate of toxicity. The values for the other chemicals are significantly higher than the experimentally determined values, as much as 3.5 log units in the case of butylamine.

The canonical correlation analysis has given a reasonable prediction of the toxicity of heptanol, ethanal, and aniline (although it must be noted that the alcohols are also well predicted by regression analysis). The predictions for

heptanol and ethanal may also be improved if parameters describing the branching of the carbon skeleton (i.e. the cluster molecular connectivities) were not included in the canonical variates. As heptanol and ethanal have a straight alkyl chains, there is no branching and consequently the cluster molecular connectivities are zero, which may affect the accuracy of the prediciton. Of the other compounds tested, only poor estimates of toxicity are obtained. However, the predictions for 4-chlorobenzaldehyde and butylamine were better than those obtained from principal component analysis.

4.2.6.2.Prediction of Fathead Minnow Toxicities

Again an interesting variation in the estimates of toxicity is obtained from the different methods (the results are listed in Table 3.16). The QSAR calculated from all the chemical classes gives some very reasonable toxicity values, with the exception of 2,4-dihydroxymethylbenzoate which is very poorly predicted, and unlike the results for the Microtox they are generally better than the results obtained from the QSARs for the separate classes. This is despite the correlation for the Microtox relationship ($r^2adj=0.661$) being slightly higher than that for the fathead minnow ($r^2adj=0.649$).

Good predictions of toxicity were found for 2-dodecanone and heptanol when the QSARs based on the individual chemical classes were used. Yet again, both of these relationships were very strong, allowing a good prediction of toxicity. A good relationship was also found for the amines, but this only revealed reasonable estimates of the fathead minnow toxicity. The calculated values of 2,4-dihydroxymethylbenzoate, acetonitrile, and 2-chlro-5-nitrobenzaldehyde were all inaccurate, even though the equation for the esters was reasonable. The relationships for the aldehydes and amines are weak and insignificant, and it is no surprise that poor estimates are obtained.

All the predictions from the regression analysis on principal components were very much higher than the true value. The reason for the repeated overestimation of toxicity by this particular statistical method is difficult to ascertain, but

may be a reflection of the fact that the principal components do not describe the important molecular features that affect a chemical's toxic action, an example being that no one principal component is based solely on hydrophobicity.
A reasonable prediction of the toxicity of heptanol is achieved from canonical correlation analysis. The remaining results are very poor, and cannot even be considered to be as accurate as CCA gave for the Microtox test. The same arguments given for error in the Microtox value are equally valid here.

4.2.6.3. An Overview of the Validity of the QSAR Techniques

Error in the predicted values of toxicity may come from a variety of sources. The greatest error in any estimate of toxicity will come from the statistical model itself. It has already been observed that with a poor, or spurious, relationship, there is little chance of obtaining an accurate toxicity value, whereas strong relationships generally lead to better predictions. Other disparities in the data may arise from the possibility of significant error in the published toxicity data. Kollig and Kitchen (1990), for instance, detail problems known to be associated with published environmental fate data. These include publishing data without reference to quality, or reliability; citations from publications which have not been substantiated; or simply misquoting numbers or other mistakes in the communication. However, with techniques as well standardised, and practised, as the two tests used for the analyses, these factors should be reduced to a minimum. Also unlikely are errors in the molecular descriptors; these are all calculated values so there can be no experimental errors. Some calculated parameters are, however, known to give unrealistic values, e.g. Bradshaw and Taylor (1989) detail complications in the structure of benzene derivatives that the ClogP algorithm does not consider, and which may lead to errors; in addition, during the molecular modelling process compounds dependent on conformation can be in error, since the true minimum energy conformation may not be found. Also there are unaccountable errors that may occur; with any data set involving manual entry of data there is always a possibility of transcription errors occuring (although every effort is made to keep these to a minimum); and molecular features that may

cause a different toxic response. Thus care must be taken in the choice of chemicals for which predictions are to be made.

Of the models tested, the greatest success came with the use of those strong relationships (i.e. having high r^2, significant F value etc.) for the individual chemical classes, and this probably reflects the likelihood of the chemicals all acting by a similar mode of toxic action. The QSAR for the toxicity of all classes of chemicals to the fathead minnow proved a reliable and accurate source of toxicity prediction, in contrast to the QSAR of all classes for the Microtox test which proved less reliable. The use of the validation technique has enabled the utility of these two relationships to be put to the test, and although they have similar statistical properties (i.e. r^2, F value etc.) only the equation for the fish toxicity has useful predictive power. Regression analyses of this kind also proved much easier to use and to comprehend than did the regression on principal components and canonical correlation analysis, as well as providing more accurate results. The results from the QSARs based on PCA and CCA were overall very disappointing, and the value of these models as predictive tools must be seriously questioned. (Although not tested, it is probable that the same problems inherent in the use of PCA would be similar for factor analysis). The poor performance of these multivariate techniques that rely on the production of new parameters from the original data may be due to the lack of definition of the new parameters, which have been unable to describe molecular features useful in the prediction of toxicity, and have so failed in their objective to help simplify and clarify the data. The other data reduction technique utilised, cluster analysis on the variables, did however succeed in reducing the dimensionality of the data matrix and by using the original variables has maintained the more pertinent features of the data.

The validation techniques have emphasised that QSAR models must be carefully considered before they can be put into practical application. It is also difficult to put limits on what may be considered as a good prediction of

toxicity. In this study a good calculated value was considered to be within approximately 0.5 log units above or below the observed value. This was an arbitarily set level, there being few guidelines to suggest whether this is a reasonable level of error to expect in a prediciton, if it is too optimistic, or even not accurate enough to be acceptable.

4.3 Inter-Species Relationships of Toxicity

A large data base of 218 comparative toxicity values has been compiled for four varied, yet commonly utilised, aquatic species. N.B. This data base is a separate, and much larger, data base from that used in the QSAR analysis but does contain some of the fathead minnow and Microtox toxicity values used to form the QSARs. Individual relationships are initially discussed, followed by a broader analysis of the methodology and results.

4.3.1.1. Relationship between Fathead Minnow and Microtox Toxicities

Only a fair relationship is found between the two toxicities ($r^2adj=0.651$; eqn 3.52); see section 3.3.1 for the full results. 38 of the 126 chemicals in the relationship have been identified as outliers (see Table 3.17), i.e. they are significantly more toxic to one or other of the species. There is a trend for reactive chemicals to be more toxic to the fathead minnow; these include those chemicals (notably the aldehydes) which can react with nucleophiles by an addition reaction, and highly substituted phenols and anilines which may act as uncouplers of oxidative phosphorylation. More toxic to Microtox tend to be those chemicals acting by polar narcosis, such as the mono-substituted phenols. Also there are a larger number of simple ketones and alcohols (which are expected to act as simple narcotics) more toxic to Microtox as compared to those more toxic to the fathead minnow. Thus it might be reasoned that the more reactive chemicals seem to be more toxic to the fish than the bacteria, which may be as a result of the difference in the length of time involved in the toxicity tests - the bacteria being exposed for 5 minutes and the fish for 96 hours. The longer test time for the fish is likely to allow chemicals to be metabolised into more toxic forms, and may also allow greater bioaccumulation to occur in the fathead minnow resulting in greater toxicity to the fathead minnow (see section 4.2.5.3 for a more detailed explanation of this matter). In addition, compounds such as the alcohols and ketones may be metabolised into less toxic compounds, and eliminated from the fathead minnow in a more efficient manner, meaning that the relative effect of baseline toxicity on the two species is different. Of the inter-species

relationships for chemical classes believed to be acting as simple narcotics, both the alcohols (eqn 3.53) and the esters (eqn 3.55) have slopes significantly less than one, indicating that there is a different relative toxic effect.

There is little pattern in the spread of the nitriles and the aliphatic amines. The nitriles, known to decompose to cyanides, are found to be relatively more toxic to both species. Malononitrile and chloroacetonitrile are the most significant outliers more toxic to the fathead minnow. In contrast 1,6-dicyanohexane is more toxic in the Microtox test, and other nitriles were not found to be outliers. The reason for increased susceptibility of the fish in comparison to the bacteria only in some cases is difficult to explain, but may be the result of more specific reactions occurring at the active site, or the individual ability of the nitriles to decompose to cyanides. Aliphatic amines are also spread throughout the data, with a tendency to be outliers to either species. For both the nitriles and amines, their unpredictability is characterised by poor inter-species relationships of toxicity (eqns 3.64 and 3.65).

Also there is no consistency with the spread of the pesticides, e.g. permethrin, an insecticide, is relatively toxic to the fathead minnow, but bromacil, a herbicide, is more toxic to Microtox. However, other insecticides such as malathion, diazinon and carbaryl were found to have similar relative toxicities. This phenomenon may arise from the very specific nature of pesticide action, which may affect any one aquatic species dramatically more than another.

The chi-squared analysis of the effect of structural features on toxicity confirms that an aldehyde group leads to greater toxicity in the fathead minnow. This reinforces the theory that such chemical classes are important in understanding the toxic effects of the compounds on these species. None of the other chemical features is significant at the 95% level.

4.3.1.2. Relationship Between Fathead Minnow and Daphnia magna Toxicities

A good relationship between the two toxicities has been found ($r^2adj=0.750$, see section 3.3.2). Again reactive chemicals, such as the alkyl halides (known to react in substitution reactions with nucleophiles), and acrolein, are more toxic to the fathead minnow. However, of the chemicals more toxic to D. magna, there are two simple alcohols, and several compounds that may act as uncouplers of oxidative phosphorylation and polar narcosis. There is also a tendency for chemicals containing an alcohol group to be more toxic to D. magna. Perhaps surprisingly D. magna is comparatively least resistant to the toxic effect of aniline (i.e. it is the greatest outlier), whereas aniline is not an outlier in the fathead minnow-Microtox correlation, and only a 'slightly' significant outlier (ratio calulated/ observed toxicity = 5.61) in the fathead minnow-T. pyriformis correlation. The reason for the invertebrate's increased susceptibility to aniline is not known.

Of the two pesticides in the data set, the predicted toxicity of permethrin is close to the observed (whilst as a selective insecticide it might have been considered to be relatively more hazardous to D. magna), whereas the toxicity of another insecticide malathion is a significant outlier being more toxic to D. magna. This again emphasises the problem of trying to predict the toxicity of such chemicals, which probably act by very specific mechanisms. The pattern of the outliers is harder to distinguish in this relationship, but again it seems likely that the effect of a longer test time enables reactive chemicals to be metabolised by the fish, whereas the effect of some simple narcotics is relatively greater to D. magna.

The inter-species relationship is improved when certain chemical classes are considered. Again a high correlation is obtained for the alcohols (eqn 3.69) and perhaps surprisingly considering their reactive nature (suggesting that they may be acting by a similar mode of action), the alkenes (eqn 3.70).

The chi-squared analysis also indicates that halogens (e.g. the alkyl halides) are more toxic to the fathead minnow, and alcohols more toxic to D. magna, in

agreement with the respective metabolism of these chemicals suggested above. Although the chi-squared analysis is not significant at the 95% level, it does show trends within the data.

4.3.1.3. Relationship Between Fathead Minnow and Tetrahymena pyriformis Toxicities

A very good relationship is found between the toxicities of these two species ($r^2adj=0.807$, see section 3.3.3). This makes it difficult to establish any pattern amongst the outliers, because there are so few (only 11 outliers out of a total of 74 compounds) and none of these has a significantly high or low ratio of calculated/observed effect to make them important to the relationship. Also it must be remembered that the T. pyriformis test has, so far, only been performed on aromatic compounds, thus the relative toxicity of such chemical classes as the simple alcohols and ketones has yet to be assessed. Of the compounds relatively more toxic to the fathead minnow, three at least (2,3,5,6-tetrachloroaniline, 1,4-dinitrobenzene, pentabromophenol) may be acting as uncouplers of oxidative phosphorylation. However all the chemicals that are more toxic to T. pyriformis (4-fluoroaniline, aniline, 4-bromoaniline, 4-chloroaniline, 4-nitrophenol) may well be metabolised to more reactive intermediates. This again is similar to the trend observed whereby compounds acting as respiratory uncouplers exert a greater relative effect on the fathead minnow.

Again an improvement can be made on the overall relationship by the consideration of correlation within chemical classes such as the ethers, ketones, pyridines, and alcohols. No structural features were statistically more prone to be outliers, but although the chi-squared value (4.81) is not significant at the 95% level, there seems to be a definite trend for anilines to be more toxic to both species. This is likely to be a consequence of their being more commonly associated with specific toxic modes of action.

4.3.1.4. Relationship Between Daphnia magna and Microtox Toxicities

A fair relationship is found (r^2adj=0.619) for all 52 compounds in this data set (see section 3.3.4). However there are 20 significant outliers. Of the chemicals more toxic to D. magna the tetrachlorophenol may act as an uncoupler of oxidative phosphorylation, the dichloroaniline may be metabolised to a more reactive form, and acrolein will react with a nucleophile, but no overall pattern emerges from the modes of toxic action. Notably more toxic to Microtox are three alkyl halides which may react with nucleophiles; it may be noted that it was these chemicals that were relatively less toxic to D. magna in the relationship with the fathhead minnow, which suggests that the invertebrate is less affected by these compounds. Several narcotics were more toxic to the Microtox, including alcohols and substituted benzenes. It is surprising that an unreactive compound such as benzene is relatively so much more toxic to Microtox than D. magna. Again no clear pattern can be elucidated. Of interest, however, should be the two pesticides present in the data (permethrin and malathion), both of which are considerably (620 and 150 times respectively) more toxic than expected to D. magna. This suggests that D. magna is considerably more susceptible to their specific modes of action, as would be expected in their role as selective insecticides. In addition, in the fathead minnow-Microtox correlation, permethrin was considerably more toxic to the fathead minnow, suggesting in this case Microtox is less sensitive to permethrin than other organisms. This may also be a factor of a longer test time enabling the pesticide to have a greater effect on both D. magna. and the fathead minnow.

Excellent correlations are found when the esters, nitrogenous coumpouds, and alkenes are considered. There are, however, relatively few data points, so these equations should be treated with caution in case a chance correlation has occured. Yet again though, a good correlation is obtained for the alcohols, for which many more data were available.

The chi-squared analysis of structural features is disappointing, showing that none of those features analysed was signifcantly more toxic to one species than

the other.

4.3.1.5. Relationship Between Daphnia magna and Tetrahymena pyriformis Toxicities

Only 17 corresponding toxicities have been obtained, revealing a poor relationship (r^2adj=0.381, see section 3.3.5). Due to the low number of data points and the poor correlation, little relevant information can be gathered from this analysis. This relationship is not significantly improved by analysis of the chemical classes present in the data set, even the alcohols giving a very unstable relationship.

4.3.1.6. Relationship Between Microtox and Tetrahymena pyriformis Toxicities

For the two microbial species, i.e. the two most closely related species of the four considered, a strong inter-species relationship might be expected. However, only a poor correlation is obtained for the 54 compounds in the relationship (r^2adj=0.317), with 12 significant outliers, (see section 3.3.5). Little pattern can be associated with these outliers, which may reflect the overall poor state of the correlation. Of the chemicals relatively more toxic to Microtox, three may be metabolised to more reactive intermediates, and another two may be acting by polar narcosis. Four of the chemicals relatively more toxic to T. pyriformis may be may be metabolised to reactive intermediates, this is similar to the T. pyriformis outliers found in the fathead minnow relationship, and may be a consequence of the longer test time for the ciliate.

The inter-species relationship is not significiantly improved by the analysis of separate chemical classes. The chi-squared analysis of structural features shows a trend (though not found to be significant) for halogenated compounds to be less likely to be more toxic to either species. This may be because only highly halogenated compounds will act by the with a more specific mode of action and thus become outliers.

4.3.2. Inter-Species Relationships - An Overview

Of the six interspecies relationships studied, the best (and possibly most interesting) were those involving the fathead minnow data, especially between the fathead minnow and Tetrahymena pyriformis toxicities (r^2adj=0.807; Eqn 3.78). Daphnia magna (r^2adj=0.750; Eqn 3.68) and Microtox (r^2adj=0.651; Eqn 3.52) toxicities are also relatively well correlated to fathead minnow toxicity. More disappointing however were the remaining three relationships concerning the prediction of Daphnia magna toxicity from Microtox toxicity (r^2adj=0.619; Eqn 3.88) and from T. pyriformis toxicity (r^2adj=0.381; Eqn 3.97), and Microtox toxicities from T. pyriformis toxicities (r^2adj=0.317; Eqn 3.101). However, the prime aim of this study was centred on the prediction of fish (fathead minnow) toxicity from a consideration of toxicities to other species.

The predictability of fathead minnow LC_{50} can be seen as a particular achievement considering the problems in the obtaining of biological data. Obtaining such data is fraught with difficulties, as errors inherent in any experimental design will affect the quality of the results. It is interesting to note that the best relationship found, that between the fathead minnow and T. pyriformis, involves data generated at only one laboratory, and by the same experimenters, for each species. Thus any interlaboratory differences in methods or protocols that may cause errors have been eliminated. It might be considered surprising therefore that the D. magna toxicity data are so well correlated to fish toxicity, when the origins of the D. magna data are taken into account - they have been taken from twelve laboratories, each using slightly different protocols. The Microtox test has however been well standardised so interlaboratory differences should be diminished; however, this has not led to such good correlations. The overall effect of the toxicity of a chemical may therefore be dependent on the length of time of the individual toxicity test. Reactive chemicals, therefore, exert a greater toxic effect after a longer time, whereas simple narcotics are unaffected by the length of exposure. The fathead minnow test has the longest exposure period (96 hours) and hence some of the chemicals will have a different toxic

effect than after five minutes for the Microtox test; consequently a lower correlation is found than for the other tests with increasing exposure time and greater correlations respectively for D. magna (48 hours) and T. pyriformis (60 hours).

The data used in the present study do not compare well with other recent studies which have revealed somewhat better correlations e.g. for the fathead minnow and T. pyriformis toxicity relationship Schultz et al (1990) found for 11 simple narcotic alcohols and ketones (it should be noted that these compounds are known to act by a single simple mechanism):

$$\log(1/LC_{50})_{FM} = 1.265\ \log(1/IG_{50})_{TP} + 0.595 \qquad (4.8)$$

n=11 s=0.200 r^2=0.988 F=743

Cajina-Quezada (1988) observed for 11 nitro- and halogen- substituted phenols and anilines:

$$\log(1/LC_{50})_{FM} = 1.254\ \log(1/IG_{50})_{TP} + 0.133 \qquad (4.9)$$

n=11 s=0.290 r^2=0.915 F=96.6

and Schultz et al (1989a) reported for 38 nitrogen containing aromatic compounds:

$$\log(1/LC_{50})_{FM} = 1.091\ \log(1/IG_{50})_{TP} + 0.284 \qquad (4.10)$$

n=38 s=0.319 r^2=0.887 F=282

The slope (0.990) and intercept (0.352) of Eqn 3.78 are similar to those in Eqns 4.8- 4.10 above, suggesting that although they are a wider range of chemical classes a similar relationship is being modelled.

The relationship between fathead minnow and Microtox toxicities has also been studied by Schultz et al (1990) who reported a good correlation for 11 simple narcotic chemicals:

$$\log(1/LC_{50})_{FM} = 0.830\ \log(1/EC_{50})_{M} + 0.004 \qquad (4.11)$$

n=11 s=0.218 r^2=0.986 F=622

and for 31 unspecified 'priority pollutants' Blum and Speece (1990) found:

$$\log(1/LC_{50})_{FM} = 0.79\ \log(1/EC_{50})_{M} + 0.60 \qquad (4.12)$$

n=31 s=0.46 r^2=0.83 F not given

Thus, again for simple narcotic toxicants the authors found a very good correlation (Eqn 4.11), yet with a more diverse data source (Eqn 4.12) the correlation is reduced. The intercept (0.189) and slope (0.704) in Eqn 3.52 are similar to those in Eqns 4.11 and 4.12. Thus, overall the equations are describing the whole relationship, and the significant outliers can be identified.

When the structural features of the compounds used in the present study are analysed, some improved relationships are observed compared to the overall equations. Some of the chemical classes have enough corresponding data to allow comparison between the individual inter-species relationships. The alcohols, for instance, are of great interest as they provide very good correlations for the prediction of fathead minnow toxicity from all three species. The slopes of these equations (0.834, 1.03, and 0.949 for Microtox, D. magna, and Tetrahymena pyriformis respectively), are all similar within the boundaries of experimental error, i.e. 0.9 $\pm$0.1, which indicate a similar toxic effect is occurring in all species. However, there are different intercepts 0.010, -0.275, and 0.425 respectively, which show that there are different sensitivities of the species to these compounds. Factors affecting the slopes and intercepts of these regression lines include the species-related differences in pharmacokinetics, such as metabolic transformation and the pharmacodynamics mediated at the level of interaction of the xenobiotic, or its metabolite with the respective receptor molecules (Wallace and Niemi, 1988). It is, unfortunately, very difficult to qualify the inter-species relationships for the alcohols in terms of the mode of toxic action, as compounds containing an -OH group may act by one of several modes; e.g. straight chain aliphatic molecules will act as simple narcotics, phenols may act as polar narcotics or uncouplers of oxidative phosphorylation depending on any substitution.

Much information is also available for nitrogenous compounds, a good relationship being found between fathead minnow and D. magna toxicities, and between D. magna and Microtox toxicities. However, other relationships for this chemical class

were found to be poorer, which may be a result of the different modes of toxic action that may occur with the various nitrogenous compounds; e.g. aliphatic nitrogenous compounds can act as simple narcotics, highly nitro-substituted aromatics can act as uncouplers of oxidative phosphorylation.

Good inter-species correlations were obtained for the alkenes between fathead minnow and both Microtox and D. magna toxcities, in addition to the Daphnia magna and Microtox relationship. The alkenes have very specific modes of action, reacting readily with nucleophile groups, so it may be considered surprising that their toxicities are well correlated. There are, however, relatively few data in these relationships (n=7, 4 and 6 respectively) so care must be taken in their interpretation.

Less satisfactory, however, were the relationships involving amine toxicity; this may be a result of amines having several specific modes of action (e.g. being metabolised to reactive intermediates, or acting as uncouplers) which may exert different toxic effects on the separate species. In addition, of the six inter-species relationships for halogenated compounds, only that for fathead minnow and D. magna has a high correlation. The halogenated compounds are a very chemically mixed group of compounds (simple halogenated alkanes and benzenes acting as narcotics, other alkyl halides may be metabolised to reactive intermediates, highly halo-substituted aromatics acting as uncouplers etc), so much so that they are probably not worth considering as a class in future studies.

In addition to the study of particular chemical classes, it would be useful to investigate groups of compounds with similar modes of action. The problem is that for many toxicants, modes of toxic action have yet to be determined, and the boundaries between different modes of action can often become a 'grey area' until more research is performed to clarify the situation. This would inevitably lead to confusion in the definition of the different groups.

4.3.2.1. Discussion of the Analysis of Outliers from the Inter-Species Relationships

There are many methods of analysing and quantifying the outliers from a regression equation, although the method used here is possibly one of the simplest. Intrinsic in its definition is that the regression line is an accurate representation of the actual relationship, and that there are likely to be roughly the same number of significant outliers on either side of the line. The outliers are defined as being outside a certain distance from the line, in this project being defined as a greater than fourfold difference between the observed and predicted values of toxicity. The distance is obviously very important, and the one chosen is the same as that used by Wallace and Niemi (1988). In addition others studies of comparative toxicity have observed that predictions can be made only with a certainty of an order-of-magnitude (i.e. $\pm$0.5 log units) (Janardan et al, 1984; Suter and Rosen, 1988). Thus the limits used in this study (a total of an eightfold scalar, or 0.9 log unit spread) seem reasonable as the limits of an accurate prediction, and compounds falling outside these limits can be considered as significant outliers. Outliers have been found ranging from 15% to 38% of the total number of data-points in a correlation, and this number is obviously related to the strength of the overall relationship. Thus this gives a rough estimate of the probability of correct prediction (within the defined limits) of a chemical's toxicity from one species to another.

An outlier may be considered to to be a chemical that has a significantly higher or lower susceptibility in one species as compared to the other. The variation in toxicity between species may be attributable to the chosen species being relatively insensitive to the action of a chemical rather than the other speices being much more sensitive (Thurston et al, 1985). The analysis of the outliers has given an indication of what may cause a chemical to be relatively more toxic to one species in a relationship. Variations in modes of toxic action are well recognised as causes of the differences in susceptibility of chemicals to different species (Maki, 1979; Thurston et al, 1985). Examination of the possible

modes of action of these chemicals shows that relatively more reactive chemicals seem to be more toxic to fathead minnow, and this may be the result of longer exposure times. (There is, of course, the problem of determining the modes of action of many of the compounds). Chemicals which react with nucleophiles by an addition reaction in the fathead minnow and Microtox study are a good example. Also the fathead minnow may be less susceptible to chemicals acting as polar narcotics, and to some simple alcohols and ketones. (Possibly it has increased ability to cope with these simple narcotics). These latter findings do not however agree with those of Thurston et al (1985) who concluded that only chemicals with specific modes of toxicity (other than narcotics) would cause large inter-species variations of toxicity because of the individual physiological processes and requirements in each species. Whilst this theory is undoubtedly valid, it takes no account of the possibility that some species may form a 'resistance' to some forms of chemical attack.

The reasons for these differences in the susceptibility of chemicals to different species are not easy to explain. In studies of the comparative susceptibilities of different aquatic species, no one species has been found to be the most susceptible overall (Suter and Rosen, 1988; Thurston et al, 1985). This may be due to several factors, such as the effect of time and metabolism to more specific toxicants (see section 4.3.1 for more details). The intrinsic diversity between species is, of course, very great (e.g. differences in their physiology and biochemistry etc.) and it must also be remembered that species metabolic responses may be a function of intra-species features such as age (or life-stage), body weight and sex (Janardan et al, 1984). In addition, Wallace and Niemi (1988) suggest that within the metabolism of a species, it may have adapted a greater enzymic metabolism for the detoxification of certain chemicals. Other more basic features of an organism could influence the relative toxicities; for example Vaishnav and Korthals (1990) consider that with microbial cells a chemical has first to move through the cell wall and often both capsule and cell

wall before reaching the cell membrane or any other protoplasmic target site. These additional barriers, capsule and cell wall, may offer some protection to microorganisms against toxic chemicals. Similarly, although not relevant to the species considered in this study, some microorganisms can rapidly produce endospores resistant to many forms of chemical attack, and with a short life-span, mutations in the genetic make-up of a microbe may occur to produce resistance.

There seems little pattern to the pesticide toxicities and which species they are more toxic against; perhaps, with specalised chemicals such as these, which obviously have highly specialised toxic actions, it is better to not consider extrapolating their biological activity. This is unfortunate, because it is such chemicals as pesticides which are of great environmental importance, having caused many pollution incidents. Janardan et al (1984) also found pesticides to be significant outliers from regression analysis of inter-species relationships of toxicity. An example is for organophosphate and carbamate pesticides (including carbaryl, diazinon, and malathion in the data in this study) which have been designed to be toxic by interacting with a specific protein, the acetylcholinesterase (AChE) molecule. The in vitro inhibitory activity of these compounds is known to vary widely among these species (Wang and Murphy, 1982) and hence will cause differences in susceptibility. In addition Maki (1979) observed that many pesticides are formulated for control of a particular target species, while creating minimal effects on non-target species. Thus differences between the susceptibility of species is only to be expected.

Overall, however, there are few clearly defined rules, because many chemicals acting by these more reactive modes are classifed as non-outliers, and indeed only one structural feature (aldehydes in the fathead minnow and Microtox relationship), is shown to cause a significant deviation from the expected numbers, as classified by the chi-squared analysis. This may, of course, show that the individual structural features are unimportant, or it may be the case that the chi-squared test is too insensitive, or even inappropriate, to study

this phenomenon. The chi-squared analysis did, however, indicate some interesting trends within the toxicity data, an example being the anilines being more prone to be outliers in the fathead minnow and T. pyriformis relationship. An earlier study of this kind (Wallace and Niemi, 1988), which did not utilise the chi-squared test however, found specific chemical classes such as aldehydes and esters caused greater toxicity to fish, whereas organophophorus insecticides were more toxic to rodents, in a fish-to-rodent inter-species comparison. Therefore the lack of structural features causing a significant increase (or decrease) in toxicity may be due to how closely the species are associated, the differences being marked in a fish-to-rodent comparison, as compared to comparisons between aquatic species.

The results analysing the outliers from the ClogP-toxicity graph show there is a definite tendency for chemicals that are outliers in the ClogP relationship to be more toxic to that organism in the interspecies relationship. This is especially true for the toxicities in the fathead minnow and Microtox, and with T. pyriformis relationships for which significant chi-squared analyses were obtained. Yet again, however, this is not a hard and fast rule, as the majority of the ClogP outliers are non-outliers from the interspecies relationships. In addition, outliers from the ClogP correlation are more toxic than expected (emphasising that they are operating by a specific mechanism), and so it should be no surprise that they are relatively more toxic to the species concerned. It should also be noted that the designation of a chemical as an outlier to the ClogP relationship does not identfy by which mode of action the chemical may be acting.

4.3.3. Use of Additional Physico-Chemical Parameters to Improve Inter-Species Relationships

Each of the inter-species relationships is improved by the addition of a physico-chemical parameter. Hydrophobicity is best correlated of the variables in all of the relationships as chosen by the stepwise regression routine (Eqns 3.66, 3.76, 3.86) and explains an additional 3-7% of the variance in the toxicity data. Obviously the toxicity data are well correlated to ClogP, and it may describe the differing effect of transport of the xenobiotics in different species. In each case another significant parameter is found. The inclusion of an indicator variable for the aldehyde group in the fathead minnow and Microtox relationship again emphasises the influence of this group. The parameters included in the other relationships are molecular connectivities, which model the bulk of a molecule. The negative correlation with first order molecular connectivity in the fathead minnow-*D. manga* relationship may be due to the limiting effect of size of a compound in passing through a membrane (this matter is discussed more fully in section 4.2.3.2). In the fathead minnow-*T. pyriformis* correlation the improvement in correlation by introducing the molecular connectivity is negligible and should be ignored. These additional parameters do not, however, increase the correlations greatly, and in each case there is a decline in the F statistic, indicating a loss in the significance of the equations.

4.3.4. The Most Representative Species

Whether it is right to quantify a species as the 'most representative' of the toxic effects of chemicals is extremely controversial (Cairns, 1986). Undoubtedly though, if few toxicity data exist, and the resources are not available to perform multi-species testing, it is essential that more information can be estimated in a cheap and reliable manner. The results show *T. pyriformis* to have the lowest positive residuals per chemical, and so to fit the definition of the most 'representative species'. The toxicity test undoubtedly benefits from a long exposure time (60 hours); in addition it is cheaper than fish tests, and the strict protocols enforced by the authors mean that it is well standardised

(Schultz et al 1990b). However, yet again it must be noted that T. pyriformis data are at present available only for aromatic compounds, and ideally some data for aliphatic compounds should be investigated before T. pyriformis could be confirmed as the most representative species.

Concluding Remarks

This study has demonstrated that it is possible, in modern QSAR, to produce a large number of physico-chemical data - how much of this is relevant and useful is open to conjecture. The data matrix has to be reduced, or condensed, before statistically valid relationships can be formed; several different approaches to this end were undertaken. Cluster analysis greatly reduced the number of variables and the redundancy between them, while retaining their original meaning. Principal component analysis, factor analysis, and canonical correlation analysis all produced new, orthogonal variables, the physico-chemical meaning of which was more difficult to establish.

In the QSAR analysis itself, regression analysis on the de-correlated variables performed much better than the use of principal components and the factors. Validation of the techniques was proved essential, as despite good statistical foundations many QSARs gave only poor predictions of toxicity. In particular those QSARs based on principal components and factors gave large amounts of error in their predictions.

The most accurate predictions of fathead minnow toxicity were given from the QSAR regression equation for all chemical classes. In contrast, for the QSARs for the Microtox data, regression analysis of separate chemical classes proved more successful. It was no surprise that hydrophobicity and steric parameters were important, especially when compounds with a similar mode of action were studied. In future studies it would, perhaps, be more profitable to study chemicals with similar modes of toxic action, as opposed to the use of separate chemical classes. The biggest disadvantage with this approach would be the correct identification of each chemical's mode of toxic action.

Some good inter-species relationships of toxicity were observed. Statistically the best were for the prediction of fathead minnow toxicities and in particular the relationship with Tetrahymena pyriformis toxicities (possibly because the T. pyriformis data were the least heterogeneous). Better correlations were often obtained when separate chemical classes, such as the alcohols, were studied.

Again, it may be more beneficial if future studies of comparative toxicity were based on chemicals with a similar mode of toxic action.

Some light has been shed on the outliers from the inter-species relationships, i.e. chemicals that exert a different relative toxic response in both species of a relationship. Much of the variation must be attributed to the intrinsic inter-species variability, affecting such basic factors as the individual's physiology and metabolism. These species-specific features are impossible to quantify fully without much more work. Another important factor is the duration of the test, which will affect the amount of detoxification, metabolism, and bioaccumulation of the toxicant.

REFERENCES

Adema D.M.M., (1978) Hydrobiologia 59:125-134.

Ahlers J., Benzig M., Gies A., Pauli W., Rosick E., (1988) Chemosphere 17:1603-1615.

Alexander M., (1981) Science 211:132-138.

Anliker R., Moser P., Poppinger D., (1988) 17:1631-1644.

ASTM (1980) Standard Practice for Conducting Acute Toxicity Tests with Fishes, Macroinvertebrates and Amphibians, ASTM Committee E-35.

Baker L.L., Wesley S.K., Schultz T.W, (1988) In Turner J.E., Williams M.W., Schultz T.W., Kwaak N.J., (eds) Proceedings of the Third International Workshop on QSAR in Environmental Toxicology. USDOE, Oak Ridge, TN, 165-168.

Balaban A.T., Motoc I., Bonchev D., Mekenyan O., (1983) In Charton M., Motoc I., (eds) Steric Effects in Drug Design. Springer-Verlag, Berlin, 21-56.

Basak S.C., Magnuson V.R., (1983) Arzneim-Forsch 33:501-503.

Basak S.C., Roy A.B., Ghosh J.J., (1980) In Avula, X.J.R., Bellman R., Luke Y.L., Rigler A.K., (eds) Proceedings of the Second International Conference on Mathematical Modelling, Vol II, University of Missouri-Rossi, MO, USA, 851-856.

Bathe R., Ullman L., Sachsse K., Hess R., (1976) Proc. Eur. Soc. Toxicol. 17:351-355.

Beckman Inc., (1981) Microtox Application Notes. M102. Advantages of Using Several Test Times.

Beckman Instruments, Inc. (1982) Microtox Systems Operating Manual.

Benoit D.A., Mattson V.R., Olson D.L., (1982) Wat. Res. 6:457-464.

Black J.A., Birge W.J., Westerman A.G., Francis P.C., (1983) Fundam. Appl. Toxicol. 3:353-358.

Blanck H., (1984) Ecol. Bull. 36:107-119.

Blum D.J.W., Speece R.E., (1990) Environ. Sci. Technol. 24:284-293.

Bobra A.M., Shiu W.Y., Mackay D., (1983) Chemosphere 12:1121-1129.

Bodor N., Gabanyi Z., Wong C-K., (1989) J. Am. Chem. Soc. 111:3783-3786.

Boethling R.S., (1986) Environ. Toxicol. Chem. 5:797-806.

Boyd D.B., Lipkowitz K.B., (1982) J. Chem. Ed. 59:269-274.

Bradbury S.P., (1988) In Turner J.E., Williams M.W., Schultz T.W., Kwaak N.J., (eds) Proceedings of the Third International Workshop on QSAR in Environmental Toxicology. USDOE, Oak Ridge, TN, 61-70.

Bradshaw J., Taylor P.J., (1989) Quant. Struct.-Act. Relat. 8:279-287.

Brooke L.T., Call D.J., Geiger D.L., Northcott C.E., (1984) Acute toxicities of

organic chemicals to fathead minnows (Pimephales promelas). Vol 1. University of Wisconsin.

Bulich A.A. (1979) In Marking L.L. and Kimerle R.A. (eds), Aquatic Toxicology, ASTM STP 667. American Society for Testing and Materials, Philadelphia, 98-106.

Bulich A.A., (1980) Interagency Workshop In-Situ Water Quality Sens. Biol. 2nd. (NTIS PB 81-164378), 173-184.

Cairns J., (1986) BioScience 36:670-672.

Cairns J., (1988) The Environ. Profess. 10:157-165.

Cajina-Quezada M., (1988) In Turner J.E., Williams M.W., Schultz T.W., Kwaak N.J., (eds) Proceedings of the Third International Workshop on QSAR in Environmental Toxicology. USDOE, Oak Ridge, TN, 123-130.

Canton J.H., Wegman R.C.C., (1983) Water Res. 17:743-747.

Chang J.C., Taylor P.B., Leach F.R., (1981) Bull. Environ. Contam. Toxicol. 26:150-156.

Charton M., (1975) J. Am. Chem. Soc. 97:1552-1556.

Clarke T., (1985) A Handbook of Computer Chemistry. John Wiley, New York.

Clayson D.B., (1988) Environ. Heal. Persp. 77:93-97.

Cololian P., (1901) J. Physiol. Pathol. (Paris) 3:535-546.

Cronin M.T.D. Dearden J.C. (1989) In Fauchere J.L. (ed) QSAR: Quantitative Structure-Activity Relationships in Drug Design. Alan R. Liss, New York, 407-410.

Cronin M.T.D., Dearden J.C., (1990) Human Expt. Toxicol. 9:118.

Cros A.F.F., (1963) Action de l'alcool amylique sur l'organisme. Thesis, Faculte de Medecine de Strasbourg, Strasbourg.

Crum Brown A., Fraser T.R., (1868) Proc. Roy. Soc. Edinburgh 25:151-203.

Curtis C., Lima A., Lozano S.J., Veith G.D., (1982) In Pearson J.G., Foster R.B., Bishop W.E., (eds) Aquatic Toxicology and Hazard Assessment: Fifth Conference, ASTM STP 766, Philadelphia, 170-178.

Dawson D.A., Lin D.T., Schultz T.W., (1988) In Turner J.E., Williams M.W., Schultz T.W., Kwaak N.J., (eds) Proceedings of the Third International Workshop on QSAR in Environmental Toxicology. USDOE, Oak Ridge, TN, 169-172.

Dayal S.K., Ehrenson S., Taft R.W., (1972) J. Am. Chem. Soc. 94:9113-9122.

Dearden J.C., (1985) Environ. Heal. Persp. 61:203-228.

Dearden J.C., (1990) In Karcher W., Devillers J., (eds) Practical Applications of QSAR in Environmental Chemistry and Toxicology. Kluwer, Amsterdam, 25-59.

Dearden J.C., Bradburne S.J.A., Cronin M.T.D., Solanki P., (1988) In Turner J.E., Williams M.W., Schultz T.W., Kwaak N.J., (eds) Proceedings of the Third International Workshop on QSAR in Environmental Toxicology. USDOE, Oak Ridge, TN, 43-50.

Dearden J.C., Bresnen G.M., (1988) Quant. Struct. -Act. Relat. 7:133-144.

Dearden J.C., Nicholson R.M., (1987) In Hadzi D., Jerman-Blazic B., (eds) QSAR in Drug Design and Toxicology. Elsevier, Amsterdam, 307-312.

De Benedetti P.G., Menziani M.C., Cocchi M., Frassineti C., (1987) Quant. Struct. -Act. Relat. 6:51-53.

de Bruijn J., Hermens J., (1990) Quant. Struct.-Act. Relat. 9:11-21.

Deneer J.W., (1988). PhD. Thesis. University of Utrecht.

Deneer J.W., Seinen W., Hermens J.L.M., (1988) Aquat. Toxicol. 12:185-192.

Deneer J.W., Sinnige T.L., Pauly J., Seinen W., Hermens J.L.M., (1987) Chemosphere 16:2365-2370.

Deneer J.W., Sinnige T.L., Seinen W., Hermens J.L.M., (1987a). Aquat. Toxicol. 10:115-129.

Devillers J., Lipnick R.L., In Karcher W., Devillers J., (eds) Practical Applications of QSAR in Environmental Chemistry and Toxicology. Kluwer, Amsterdam, 181-196.

Dewar Research Group (1986) AMPAC program, QCPE 506.

DeZwart D., Sloof W. (1983) Aquat. Toxicol. 4:129-138.

Dill D.C., Mayes M.A., Mendoza C.G., Boggs G.U., Emmitte J.A., (1982) In Pearson J.G., Foster R.B., Bishop W.E., (eds) Aquatic Toxicology and Hazard Assessment: fifth conference, ASTM STP 766, 245-256.

Drummond R.A., Russom C.L., Geiger D.L., Defoe D.L., (1986) In Poston T.M. and Purdy R. (eds) Aquatic Toxicology and Environmental Fate. Vol IX. ASTM STP 921, Philadelphia, PA, 415-435.

Dunn W.J., (1988) Toxicol. Let. 43:277-283.

Dutka B.J. Kwan K.K. (1981). Bull. Environ. Contam. Toxicol. 27:753-757.

Dutka B.J. Kwan K.K. (1982). J. Env. Poll. Ser. A. 29:125-134.

Dutka B.J. Kwan K.K. (1984). In Liu D. and Dutka B.J.J. (eds) Toxicity Screening Using Bacterial Systems. M. Dekker Inc., N.Y., 125-138.

Eastmond D.A., Booth G.M., Lee M.L., (1984) Arch. Environ. Contam. Toxicol. 13:105-111.

Edward J.T., (1970) J. Chem. Ed. 47:261-270.

Ehrenson S., Brownlee R.T.C., Taft R.W., (1973) Prog. Phys. Org. Chem. 10:1-80.

Elnabarawy M.T., Robideau R.R., Beach S.A., (1988) Tox. Assess. 3:361-370.

Enslein K., (1988) Toxicol. Ind. Heal. 4:479-497.

Enslein K., Tuzzeo T.M., Borgstedt H.H., Blake B.W., Hart J.B., (1987) In Kaiser K.L.E. (ed) QSAR in Environmental Toxicology - II. D. Reidel, Dordrecht, 91-106.

Finney D.J. (1980) Statistics for Biologists. Chapman and Hall. London.

Fisher A.N., Brown K., Davis S.S., Parr G.D., Smith D.A., (1987) J. Pharm. Pharmacol. 39:357-362.

Free S.M., Wilson J.W., (1964) J. Med. Chem. 7:395-399.

Freireich E.J., Gehan E.A., Rall D.P., Schmidt L.H., Skipper H.E., (1966) Cancer Chemother. Rep. 50:219-244.

Ford M.G., Greenwood R., Turner C.H., Hudson B., Livingstone D.J., (1989) Pestic. Sci. 27:305-326.

Geiger D.L., Call D.J., Brooke L.T., (1988) Acute toxicities of organic chemicals to fathead minnows (Pimephales promelas). Vol 4. University of Wisconsin.

Geiger D.L., Northcott C.E., Call D.J., Brooke L.T., (1985) Acute toxicities of organic chemicals to fathead minnows (Pimephales promelas). Vol 2. University of Wisconsin.

Geiger D.L., Poirier S.H., Brooke L.T., Call D.J., (1986) Acute toxicities of organic chemicals to fathead minnows (Pimephales promelas). Vol 3. University of Wisconsin.

Gersich F.M., Blanchard F.A., Applegath S.L., Park C.N., (1986) Arch. Environ. Contam. Toxicol. 15:741-749.

Geyer H.J., Scheunert I., Korte F., (1987) Chemosphere 16:239-252.

Glen R.C., Rose V.S., (1987) J. Mol. Graph. 5:79-86.

Goldberg L., (1983) Structure-Activity Correlations as a Predictive Tool in Toxicology. Hemisphere, New York.

Gough K.M., Kaiser K.L.E., (1988) In Turner J.E., Williams M.W., Schultz T.W., Kwaak N.J., (eds) Proceedings of the Third International Workshop on QSAR in Environmental Toxicology. USDOE, Oak Ridge, TN, 111-123.

Gray H.L., Schucany W.R., (1972) The Generalized Jackknife. Marcel Dekker, New York.

Gruber C., Buss V., (1989) Chemosphere 19:1595-1609.

Hall L.H., Kier L.B., (1989) Environ. Toxicol. Chem. 8:19-24.

Hammett L.P., (1935) Chem. Rev. 17:125-136.

Hansch C., Dunn W.J., (1972) J. Pharm. Sci. 61:1-19.

Hansch C., Fujita T., (1964) J. Am. Chem. Soc. 86:1616-1626.

Hansch C., Leo A., (1979) Substituent Constants for Correlation Analysis in Chemistry and Biology. Wiley, New York.

Hansch C., Maloney P.P., Fujita T., Muir R.M., (1962) Nature 194:178-180.

Hansch C., Muir R.M., Fujita T., Maloney P.P., Geiger F., Streich M.J., (1963) J. Am. Chem. Soc. 85:2817-2824.

Harris R.J., (1975) A Primer to Multivariate Statistics. Academic, New York.,

Hauk A., Schramm K.W., (1990) Toxicol. Environ. Chem. 26:45-54.

Hermens J.L.M., (1989) In Hutzinger O., (ed) Handbook of Environmental Chemistry vol 2E, Reactions and Processes. Springer Verlag, Berlin, 111-162.

Hermens J., Busser F., Leeuwangh P., Musch A., (1985) Ecotoxicol. Environ. Saf 9:17-25.

Hermens J., Busser F., Leeuwanch P., Musch A., (1985b) Toxicol. Environ. Chem. 9:219-236.

Hermens J., Canton H., Steyger N., Wegman R., (1984) Aquat. Toxicol. 5:315-322.

Hermens J., Leeuwangh P., Musch A., (1984b) Ecotoxicol. Environ. Saf. 8:388-394.

Hodson P.V., (1985) J. Appl. Toxicol. 5:220-226.

Holcombe G.W., Phipps G.L., Fiant J.T., (1983) Ecotoxicol. Environ. Saf. 7:400-409.

Holcombe G.W., Phipps G.L., Sulaiman A.H., Hoffman A.D., (1987) Arch. Environ. Contam. Toxicol. 16:697-710.

Hopfinger A.J., (1985) J. Med. Chem. 28:1133-1139.

Hunter R.S., (1988) In Turner J.E., Williams M.W., Schultz T.W., Kwaak N.J., (eds) Proceedings of the Third International Workshop on QSAR in Environmental Toxicology. USDOE, Oak Ridge, TN, 33-42.

Hyde R.M., Livingstone D.J., (1988) J. Comp. -Aid. Mol. Des. 2:145-155.

Isnard P., Lambert S., (1988) Chemosphere 17:21-34.

Janardan S.R., Olson C.S., Schaeffer D.J., (1984) Ecotoxicol. Environ. Saf. 8:531-539.

Kaiser K.L.E., Ribo J.M., (1988) Tox. Assess. 3:195-237.

Kaiser K.L.E., Ribo J.M., (1985) In Tichy M. (ed) QSAR in Toxicology and Xenobiochemistry, Elsevier, Holland, 27-38.

Kamlet M.J., Doherty R.M., Carr P.W., Mackay D., Abraham M.H., Taft R.W., (1988) Environ. Sci. Technol. 22:503-509.

Kamlet M.J., Doherty R.M., Veith G.D., Taft R.W., Abraham M.H. (1986) Environ. Sci. Technol. 20:690-695.

Kenaga E.E., (1972) Residue Rev. 44:73-113.

Kenaga E.E., (1982) Environ. Toxicol. Chem. 1:347-358.

Kerler F., Schonherr J., (1988) Arch. Environ. Contam. Toxicol. 17:1-6.

Kier L.B., (1985) Quant. Struct. -Act. Relat. 4:109-116.

Kier L.B., (1986a) Quant. Struct. -Act. Relat. 5:1-6.

Kier L.B., (1986b) Quant. Struct. -Act. Relat. 5:7-11.

Kier L.B., (1987) Quant. Struct. -Act. Relat. 6:8-12.

Kier L.B., (1989) Quant. Struct. -Act. Relat. 8:221-223.

Kier L.B., Hall L.H., (1976) Molecular Connectivity in Chemistry and Drug Research. Academic Press, New York.

Kier L.B., Hall L.H., (1986) Molecular Connectivity in Structure-Activity Analysis. Research Studies Press, Letchworth, Herts, England.

Kikuchi O., (1987) Quant. Struct. -Act. Relat. 6:179-184.

Klein W., Kordel W., Klein A.W., Kuhnen-Clausen D., Weiss M., (1988) Chemosphere 17:1445-1462.

Klopman G., (1985) Environ. Heal. Persp. 61:269-274.

Koch R., (1983) Toxicol. Environ. Chem. 6:87-96.

Kollig H.P., Kitchens B.E., (1990) Toxicol. Environ. Chem. 28:95-103.

Konemann H., (1981) Toxicology 19:209-221.

Kopperman H.L., Carlson R.M., Caple R., (1974) Chem.-Biol. Interactions 9:245-251.

Leahy D.E., (1986) J. Pharm. Sci. 75:629-636.

LeBlanc G.A., (1980) Bull. Environm. Contam. Toxicol. 24:684-691.

LeBlanc G.A., (1984) Environ. Toxicol. Chem. 3:47-60.

Leegwater D.C., (1989) Aquat. Toxicol. 15:157-168.

Leo A., Jow P.Y.C., Silipo C., Hansch C., (1975) J. Med. Chem. 18:865-868.

Lindeman R.H., Merenda P.F., Gold R.Z., (1980) Introduction to Bivariate and Mutivariate Analysis. Scott, Foresman & Co., Glenview, Illinois, USA.

Lipnick R.L., (1989) Environ. Toxicol. Chem. 8:1-12.

Lipnick R.L., Dunn W.J., (1983) In Dearden J.C. (ed) Quantitative Approaches to Drug Design. Elsvier, Amsterdam, 265-266

Lipnick R.L., Watson K.R., Strausz A.K., (1987) Xenobiotica 17:1011-1025.

Livingstone D.J., (1989) Pestic. Sci. 27:287-304.

Livingstone D.J., Rahr E., (1989) Quant. Struct. -Act. Relat. 8:103-108.

Ljublina E.L., Filov A.V., (1975) In: Methods used in the USSR for Establishing Biologically Safe Levels of Toxic Substances. W.H.O. Geneva, 19-43.

Maki A.W., (1979) J. Fish. Res. Board Can. 36:411-421.

Manly B.F.J., (1986) Multivariate Statistical Analysis. A Primer. Chapman and

Hall, London.

Martin Y.C., (1978) Quantitative Drug Design. Marcel Dekker, New York.

McCarty L.S., (1986) Environ. Toxicol. Chem. 5:1071-1080.

McCarty L.S., Hodson P.V., Craig G.R., Kaiser K.L.E., (1985) Environ. Toxicol. Chem. 4:595-606.

McFeters G.A, Bond P.J., Olson S.B., Tchan Y.T., (1983) Wat. Res. 17:1757-1762.

Medchem Software Manual (1987) Release 3.52. Medicinal Chemistry Project, Pomona College, Claremont, CA.

Metcalf R.L., Sanborn J.R., Lu P-Y., Nye D., (1975) Arch. Environ. Contam. Toxicol. 3:151-165.

Meyer H., (1899) Naunyn-Schmiedeberg's Archiv fuer Experimentelle Pathologie und Pharmakologie. 42:109-118.

Moulton M.P., (1988) In Turner J.E., Williams M.W., Schultz T.W., Kwaak N.J., (eds) Proceedings of the Third International Workshop on QSAR in Environmental Toxicology. USDOE, Oak Ridge, TN, 1-10.

Murray W.J., (1977) J. Pharm. Sci. 66:1352-1354.

Neely W.B., Branson D.R., Blau G.E., (1974) Environ. Sci. Technol. 8:1113-1115.

Niemi G.J., (1990) In Karcher W., Devillers J., (eds) Practical Applications of QSAR in Environmental Chemistry and Toxicology. Kluwer, Amsterdam, 153-169.

Nys G.C., Rekker R.F., (1973) Chim. Ther. 8:521-535.

Overton E., (1899) Vierteljahrsschrift der Naturforschenden Gesellschaft in Zuerich. 44:88-135.

Panthananickal A., Hansch C., Leo A., Quinn F.R., (1978) J. Med. Chem. 21:16-26.

Paris D.F., Wolfe N.L., Steen W.C., (1984) Appl. Environ. Microbiol. 47:7-11.

Paris D.F., Wolfe N.L., Steen W.C., Baughman G.L., (1983) Appl. Environ. Microbiol. 45:1153-1155.

Parsons J.R., Govers H.A.J., (1990) Ecotox. Environ. Saf. 19:212-227.

Pearlman R.S., (1981) SAREA program, QCPE 413.

Pearson E.S., Hartley H.O., (1972) Biometrika Tables for Statisticians. Vol II. Cambridge University Press, London.

Pitter P., (1985) Acta Hydrochim. Hydrobiol. 13:453-460.

Pople J.A., Beveridge D.L., (1970) Approximate Molecular Orbital Theory, McGraw-Hill Inc., New York.

Protic M., Sabljic A., (1989) Aquat. Toxcicol. 14:47-64.

Purcell W.P., Bass G.E., Clayton J.M., (1973) Strategy of Drug Design: A Guide to Biological Activity. Wiley Interscience, New York.

Purdy R., (1988) In Turner J.E., Williams M.W., Schultz T.W., Kwaak N.J., (eds) Proceedings of the Third International Workshop on QSAR in Environmental Toxicology. USDOE, Oak Ridge, TN, 99-110.

Qureshi A.A., Flood K.W., Thompson S.R., Janhurst S.M., Innis C.S., Rokosh D.A., (1982) In Pearson J.G., Foster R.B., Bishop W.E., (eds) Aqatic Toxicity and Hazard Assessment: Fifth Conference, ASTM STP 766. American Society for Testing and Materials, Philadelphia, 179-195.

Randic M., (1975) J. Am. Chem. Soc. 97:6609-6615.

Rekker R.F., de Kort H.M., (1979) Eur. J. Med. Chem.-Chim. Ther. 14:479-488.

Ribo J.M., Kaiser K.L.E., (1983) Chemosphere 12:1421-1442.

Ribo J.M., Kaiser K.L.E., (1984) In Kaiser K.L.E. (ed) QSAR in Environmental Toxicology - II. D. Reidel, Dordrecht, Holland, 319-336.

Ribo J.M., Kaiser K.L.E., (1987) Tox. Assess. 2:305-323.

Ribo J.M., Rogers F., (1990) Tox. Assess. 5:135-152.

Richards W.G., (1983) Quantum Pharmacology. 2nd ed. Butterworth, London.

Richter J.E., Peterson S.F., Kleiner C.F, (1983) Arch. Environ. Contam. Toxicol. 12:679-684.

Roberts D.W., (1987) In Kaiser K.L.E. (ed) QSAR in Environmental Toxicology - II. D. Reidel, Dordrecht, 295-308.

Ross P.E., Henebry M.S., (1989) Tox. Assess. 4:1-21.

Sabljic A., (1989) In Fauchere J.W., (ed) QSAR: Quantitative Structure-Activity Relationships in Drug Design. Alan R. Liss, New York, 411-415.

Sabljic A., (1990) In Karcher W., Devillers J., (eds) Practical Applications of QSAR in Environmental Chemistry and Toxicology. Kluwer, Amsterdam, 61-82.

Sabljic A., Protic M., (1982) Chem.-Biol. Interact. 42:301-310.

Samak Q.M., Noiseux R., (1981) Can. Tech. Rep. Fish. Aquat. Sci. 990:288-308.

Schultz T.W., (1983) In Nriagu J.O. (ed) Aquatic Toxicology. John Wiley and Sons, New York.

Schultz T.W., (1988) In Turner J.E., Williams M.W., Schultz T.W., Kwaak N.J., (eds) Proceedings of the Third International Workshop on QSAR in Environmental Toxicology. USDOE, Oak Ridge, TN, 51-60.

Schultz T.W., Dawson D.A., Lin D.T., (1989) Chemosphere 18:2283-2291.

Schultz T.W., Holcombe G.W., Phipps G.L., (1986) Ecotox. Environ. Saf. 12:146-153.

Schultz T.W, Lin D.T., Wilke T.S., Arnold L.M., (1990b) In Karcher W., Devillers J., (eds) Practical Applications of QSAR in Environmental Chemistry and Toxicology. Kluwer, Amsterdam, 241-260.

Schultz T.W., Wesley S.K., Baker L.L., (1989b) Bull. Environ. Contam. Toxicol. 43:192-198.

Schultz T.W., Wyatt N.L., Lin D.T., (1990) Bull. Environ. Contam. Toxicol. 44:67-72.

Schuurmann G., Klein W., (1988) Chemosphere 8:1551-1574.

Serat W.F., Budinger F.E., Mueller P.K. (1965) J. Bact. 90:832-833.

Serat W.F., Kyono J., Mueller P.K. (1969) Atmos. Env. 3:307-309.

Slooff W., Canton J.H., Hermens J.L.M., (1983) Aquat. Toxicol. 4:113-128.

Smissaert H.R., Jansen A.A.M., (1984) Ecotox. Environ. Saf. 8:294-302.

Smith G.H., (1986) MOLSV program, QCPE 509.

Surowitz K.G., Burke B.E., Pfister R.M., (1987) Toxic. Assess. 2:17-27.

Suter G.W., Rosen A.E., (1988) Environ. Sci. Technol. 22:548-556.

Swain C.G., Lupton E.C., (1968) J. Am. Chem. Soc. 90:4328-4337.

Szydlo R.M., Ford M.G., Greenwood R., Salt D.W., (1984) In Kuchar M. (ed) QSAR in Design of Bioactive Compounds. J.R. Prous, Barcelona, 301-320.

Taft R.W., (1952) J. Am. Chem. Soc. 74:3120-3128.

Taft R.W., (1953) J. Am. Chem. Soc. 75:4231-4238.

Tanii H., Hashimoto K., (1985) Arch. Toxicol. 55:47-54.

Tardiff R.G., Rodricks J.V., (1987) Toxic Substances and Human Risk. Plenum Press, New York.

Tichy M., Trcka V., Roth Z., Krivucova M., (1985) Environ. Heal. Perspec. 61:321-328.

Thurston R.V., Gilfoil T.A., Meyn E.L., Zajdel R.K., Aoki T.I., Veith G.D., (1985) Wat. Res. 19:1145-1155.

Topliss J.G., Costello R.J., (1972) J. Med. Chem. 15:1066-1068.

Tosato M.L., Marchini S., Passerini L., Pino A., Eriksson L., Lindgren F., Hellberg S., Jonsson J., Sjostrom M., Skagerberg B., Wold S., (1990) Environ. Toxicol. Chem. 9:265-277.

True C.J., Heyward A.A., (1990) Tox. Assess. 5:29-45.

Turner L., Choplin F., Dugard P., Hermens J., Jaeckh R., Marsmann M., Roberts D., (1987) Toxic. in Vitro. 1:143-171.

Tute M.S. (1990) In Ramsden C.A. (ed) Comprehensive Medicinal Chemistry. Vol 4. Pergamon, Oxford, 1-31.

Vaishnav D.D., Korthals E.T., (1990) Arch. Environ. Contam. Toxicol. 19:624-628

Veith G.D., Broderius J.S., (1987) In Kaiser K.L.E. (ed) QSAR in Environmental

Toxicology - II. D. Reidel, Dordrecht, 385-392.

Veith G.D., Call D.J., Brooke L.T., (1983) Can. J. Fish. Aquat. Sci., 40:743-748.

Veith G.D., De Foe D., Knuth M., (1985) Drug Met. Rev. 15:1295-1303.

Veith G.D., Greenwood B., Hunter R.S., Niemi G.J., Regal R.R., (1988) Chemosphere 17:1617-1630.

Verloop A., Hoogenstraten W., Tipker J., (1976) In Ariens E.J., (ed) Drug Design vol VII. Academic Press, New York. 165-208.

Vinter J.G., Davis A., Saunders M.R., (1987) J. Com.-Aided Design. 1:31-51.

Wallace K.B., Niemi G.J., (1988) Environ. Toxicol. Chem. 7:201-212.

Wang C., Murphy S.D., (1982) Toxicol. Appl. Pharmacol. 66:409-419.

Warne M. St.J., Connell D.W., Hawker D.W., (1989) Chemosphere 19:1113-1128.

Weininger D., Weininger A., Weininger J.L, (1986) Chem. Des. Auto. News 1:2-15.

Weiss M., Kordel W., Kuhnen-Clausen D., Lange A., Klein W., (1988) Chemosphere 17:1419-1443.

White D.N.J., (1977) Computers and Chemistry. 1:225-233.

Wold S., Dunn W.J., (1983) J. Chem. Inf. Comput. Sci. 23:6-13.

Appendix 1. The Microtox Bioassay

The use of microbial organisms as test species for environmental pollutants has been widely investigated. Microbial tests have been widely used in toxicity screening procedures due to several factors, such as similarity of complex biochemical functions with higher organisms, ease of handling, short exposure time, and reproducibility of the results between laboratories. A variety of tests have been studied, based on the measurement of different indicators of the biological status of these organisms. Enzymatic activity, growth inhibition, reproduction rate, oxygen demand, metabolic light and heat release, have all been used as measures of the toxic effects of industrial wastes and single contaminants (Ribo and Kaiser, 1987).

The measurement of the emission of light by luminescent bacteria has been used to develop a sensitive test for the quick assessment of aquatic toxicity since the first experiments in which the effect of air pollutants on such bacteria was determined (Serat et al., 1965, 1969). Several strains of luminescent bacteria and culture media were studied and evaluated, and a specific test for the rapid assessment of the toxicity of aquatic samples using the light emitting bacterium Photobacterium phosphoreum was proposed by Bulich (1979). Bulich and coworkers also developed a lyophilization (freeze drying) procedure to standardise the bacterial culture. This system has been developed commercially under the tradename Microtox, by Beckman Instruments, Inc. (1982).

The P. phosphoreum light emission spectrum ranges from 420 to 630 nm with an intensity maximum at 490nm, i.e. in the visible region of the spectrum. Intensity of the light output depends on several external factors including temperature, pH, salinity, as well as the nature and concentration of the toxicant. In order to minimise the variability between measurements, rigorous control of such external factors is necessary.

As a marine bacterium, P. phosphoreum is naturally adapted to a marine environment. The addition of NaCl to saline concentrations of 20 g/l is recommended, although the bacterium will withstand a range from 5 to 50 g/l. The

pH range for optimal physiological conditions is between 5 and 9; again the recommended range is much narrower, from 6.5-7, for test conditions. It will survive a range of temperatures from 10°C to 25°C, with light output varying appropriately. The temperature recommended and most commonly used is 15°C.

Comparisons with Other Tests

Research on the recently developed bioassay involved independent studies on its sensitivity, accuracy and precision. Evaluation of the results obtained and comparison with other bioassays have been peformed (Bulich, 1980; Dutka and Kwan, 1981; Qureshi et al., 1982; Ribo and Kaiser, 1983). Investigation centred on the applicability of the photoluminescent bacterium bioassay to different areas where quick toxicity tests were most needed, including monitoring of the toxicity of water treatment plant effluents, industrial discharges and for use as a standard toxicity indicator for single chemical compounds. The bioassay was found to compare favourably with other tests in sensitivity and reproducibility, and good correlation with other bioassays has been found for several classes of common contaminants.

The Microtox test has been compared with other microbial tests. The response of light emitting bacteria to common toxicants has been compared to the response obtained using nonluminescent bacteria such as Spirillum volutans, Pseudomonas fluorescens, Aeromonas hydrophila (Dutka and Kwan, 1981, 1982), Bacillus subtilis, and Bacillus sp. (Ribo and Kaiser, 1983). Comparisons of the Microtox toxicity with inhibition of respiratory activity of activated sludge, and inhibition of activated sludge TTC dehydrogenase activity have also been reported (Dutka and Kwan, 1981; Dutka and Kwan, 1984). More recently the Microtox test has been compared with a new bacterial test, Polytox, a new commercial preparation produced by the Polybac Corporation, 1986, utilising a specialised blend of bacterial cultures. Elnabarawy et al., (1988) found the Microtox system more sensitive to both organic and inorganic chemicals than was the Polytox system. Overall the Microtox test has been found to be rapid and simple, and to have

equal or greater precision and/or sensitivity than traditional fish bioassays. The Microtox test is also used for screening the ecotoxicological hazards of contaminated sediments and effluents. Ross and Henebry (1989) report that Microtox test results, along with three other microbial tests, agreed with community bioassays. True and Hayward (1990) found that the test had the potential to estimate both the water-soluble and solvent-soluble compounds in marine sediments.

The Use of Microtox Toxicity Data in QSAR Analysis

The Microtox bioassay has also been found very valuable in generating toxicity data to be used for the assessment and prediction of the aquatic toxicity of single chemicals through mathematical models using quantitative structure-activity relationships (QSAR). Ribo and Kaiser (1984) showed that the Microtox toxicity of chlorinated aromatic compounds could be explained and predicted using physico-chemical properties of the compounds, primarily the octanol/water partition coefficient, and other structural properties including the van der Waals volume, molecular symmetry, and hydrophilic character. More recently, Kamlet et al (1986) investigated the relationship of solute properties, expressed in terms of solvatochromic parameters, to the Microtox toxicity of nonelectrolyte organic chemicals, and found that the toxicity could be well predicted by a generalised linear solvation energy relationship. Also Gough and Kaiser (1988) discovered good correlations between the Microtox toxicity of 4-substituted nitrobenzenes and anilines and calculated electronic properties, such as the variation in the charge density at the oxygen atom of the nitro group. Ribo and Rogers (1990) have proposed an mathematical algorithm for the assessment of mixtures of chemicals.

Many interspecies relationships have been noted between Microtox toxicity and that to other aquatic species (Ribo and Kaiser, 1983; Cronin and Dearden, 1988).

Appendix 2. Molecular Connectivities

The most successful of all topological indices are the molecular connectivities. They have found numerous applications in various areas of physics, chemistry, biology and drug design, as well as the environmental sciences. Their success can be accounted for primarily in two ways:

i) They are based on sound chemical, structural (topologic and geometric), and mathematical grounds.

ii) They were developed with the idea to parallel important physico-chemical properties like boiling points, chromatographic retention times, enthalpies of formation and total molecular surface area.

These topological indices originate from the work of Randic (1975), it was however, Kier and Hall (1976) who greatly extended his work and first applied the term 'molecular connectivity'.

Calculation

In the simplest form of the index, the structure of the molecule for which information is required is expressed as a hydrogen suppressed graph. Each carbon atom is designated by a number (delta, d) which is a count of the number of adjacent (or formally bonded) carbon atoms. The molecular skeleton is dissected into all constituent bonds, each designated by the two carbon atoms i and j forming the bond. Using the Randic algorithm

$$(d_i\ d_j)^{-0.5}$$

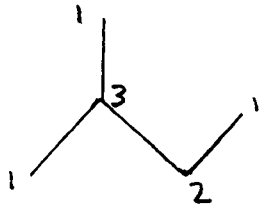
a value for each bond can be calculated. The molecular index is simply the sum of these bond values over the entire molecule, i.e.

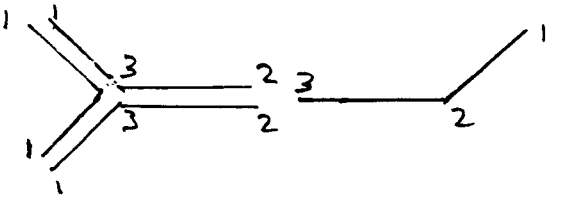
$$^1\chi = \sum (d_i\ d_j)^{-0.5}$$

It is axiomatic that a single index will not encode sufficient information about molecular structure to approximate all its complexity. Accordingly Kier and Hall (1976) proposed a scheme whereby higher order dissection of the molecular skeleton became the basis of additional extended indices.

For second order connectivity, the molecular skeleton is dissected into 'two contiguous bond' fragments, in which the delta values are maintained. Thus for

isopentane the two-bond dissection is as follows:





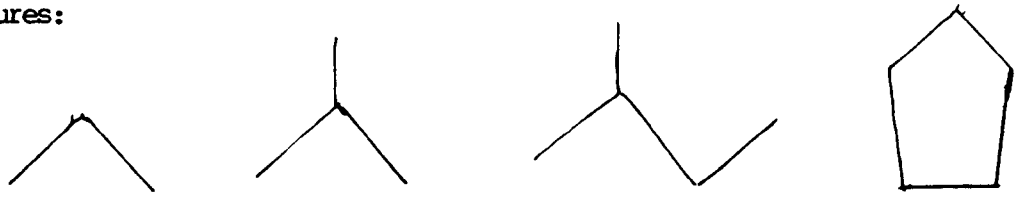
The Randic algorithm is modified to

$$(d_i\ d_j\ d_k)^{-0.5}$$

and thus the calculation of second order molecular connectivity is achieved by

$$^2X = \sum (d_i\ d_j\ d_k)^{-0.5}$$

The calculation of higher order indices is possible by dissection of any molecule into the appropriate fragments. In addition to these so-called path fragments, information can be derived by the dissection of a molecule into these other features:



a b c d

where a is path (2nd order)
b is cluster (3rd order)
c is path/cluster (4th order)
d is chain (5th order)

Obtaining the index for each of these features follows the same scheme as before. In addition, a zero order molecular connectivity can be calculated if each atom is considered as a fragment, thus the algorithm for each atom simply becomes:

$$d^{-0.5}$$

and zero order connectivity is given as (Kier and Hall, 1986)

$$^0X = \sum (d)^{-0.5}$$

Valence-Correction

Kier and Hall (1976) put forward a new rationale so that more information inherent in such features as unsaturated bonds could be formalised. This is

performed by assigning the delta value as a count of each bond to an adjacent atom. A double bond thus counts twice when adding up adjacent atoms. The hybridisation state of the carbon atoms are therefore accounted for.

The modified delta value is referred to as valence delta (d^V). the calculation then proceeds as before with valence corrected delta replacing simple delta.

Valence delta can be expressed as:

$$d^V = Z^V - h$$

where Z^V is the number of valence electrons
h is the number of hydrogen atoms suppressed

Treatment of Heteroatoms

Kier and Hall (1976) describe the treatment of heteroatoms, which is based on the explicit count of adjacent bonded atoms (excluding hydrogen) plus a count of all pi and lone pair electrons. However, when high row atoms are considered (e.g. P, S, Cl, Br, and I) specific account must be taken of non-valence, or core, electrons. These electrons play a strong and direct role in the size of atoms, and indirectly influence such properties as ionisation potential and electron affinity. The valence delta value for all heteroatoms is thus

$$d^V = (Z^V - h) / (Z - Z^V - 1)$$

Nomenclature

The Greek letter chi (symbolised as X) has been adopted (Kier and Hall, 1976) to represent the index. Two superscripts and one subscript are used to specify the particular index. The left hand superscript gives the order of the index. The right hand superscript differentiates between valence and non-valence indices. The right hand subscript specifies the subclass of index (i.e. path, cluster, path/cluster, or chain).

Physical Significance of Molecular Connectivity

As they stand, molecular connectivities tell us nothing of the physical properties of a molecule. Much work has been performed to assign some physical basis to them, and it is not surprising to find that the indices are well correlated with measures of bulk volume.

For instance Murray (1977) found that the Taft steric constant (E_S) correlated well with molecular connectivities for a series of alkyl esters:

$E_S = -0.544\ {}^2X - 1.40\ {}^3X + 1.09\ {}^4X + 0.403$

n=19 s=0.460 r^2=0.924 F not given

Kier and Hall (1986) found the following relationship for the molar volume (MV) of 37 alkanes of varying chain length:

$MV = 24.9\ {}^1X + 11.9\ {}^2X - 2.84\ {}^4X + 39.8$

n=37 s=1.17 r=0.99 F=86600

Kier and Hall (1986) also observed an excellent equation for molar refractivity (MR):

$MR = 3.83\ {}^0X + 4.44\ {}^1X - 0.873\ {}^3X - 0.483\ {}^4X - 0.456$

n=55 s=0.043 r=0.99 F=195000

In addition to these examples Kier and Hall (1986) also report many other relationships of molecular connectivities with physical properties such as polarisability, water solubility, chromatographic retention indices, and thermodynamic properties including heats of atomisation and vapourisation.

Dearden et al (1988) extended the study of physical relevance by studying the relationship of the 54 parameters listed by van de Waterbeemd and Testa (1987) for 59 substituents of varying nature. The conclusion was again that path connectivity terms, whether simple or valence, predominantly model bulk volume, including van der Waals volumes, parachor and molar refractivity.

Veith et al (1988) used molecular connectivities to explore the intrinsic dimensionality of chemical structure space. For each of 19,972 chemicals 90 molecular connectivity indices were calculated. The whole data matrix was subjected to principal component analysis. Eight significant principal components were formed that explained 93.6% of the variation of the data. These principal components were interpreted as encoding the size of molecules, and different features of its branching.

Application of Molecular Connectivities to QSAR Studies

Reports of the applicability of molecular connectivities as parameters for use in QSAR studies abound (Kier and Hall 1976, 1986; Sabljic 1990). It is probably true to say that the indices have been most extensively utilised in the field of environmental QSAR. Some examples are given below.

Hall and Kier (1989) observed that third order valence path molecular connectivity produces a good model for the LC_{50} of substituted phenols to the fathead minnow:

$$\log(1/LC_{50}) = 1.08\ {}^{3}X^{v} + 2.52$$

n=25 s=0.35 r=0.903 F=101

Leegwater (1989) repeated some of the QSAR analysis of Konemann (1981) and Hermens et al (1984b), revealing the following equations:

$$\log(1/LC_{50}) = 1.15(0.07)\ {}^{2}X^{v} - 4.25$$

n=12 s=0.17 r=0.98 F=279

where $\log(1/LC_{50})$ is the toxicity of chlorobenzene derivatives to the guppy (from Koneman, 1981)

$$\log(1/LC_{50}) = 1.46(0.16)\ {}^{2}X^{v} - 5.18$$

n=17 s=0.33 r=0.92 F=79

where $\log(1/LC_{50})$ is the toxicity of aniline derivatives to the guppy (from Hermens et al, 1984b)

In both cases there is no significant loss in the goodness of fit of the equations by using ${}^{2}X^{v}$ in preference to log P.

In other areas of environmental QSAR, Sabljic and Protic (1982) found good correlations for the bioconcentration factor (BCF) of chlorinated benzenes, PCBs, and chlorinated diphenyl oxides in fish:

$$\log BCF = 2.22\ {}^{2}X^{v} - 0.17\ ({}^{2}X^{v})^{2} - 2.32$$

n=20 s=0.277 r^2(adj)=0.936 F=139

In a related area, strong relationships have been found for modelling soil sorption of chemicals (Sabljic, 1989).

Kappa Indices

The kappa (or molecular graph shape) index is a numerical index of molecular shape derived from the graph of the non-hydrogen molecular skeleton. It is defined by Kier (1985, 1986a, 1986b) and calculated as follows. The count of the atoms A, and the number of paths of length one (^1p) in the hydrogen-suppressed graph or skeleton structure are obtained. From the relationship of p to the maximum and minimum values of p for that number of atoms, a general expression is derived (Kier, 1985):

$$^1k = A(A - 1)^2 / (^1p)^2$$

where ^1k is the first order kappa index

To account for heteroatoms, or carbon atoms other than Csp^3, a modification of each atom count in arriving at A is made by using $1 + \alpha$ for each atom other than Csp^3 (Kier, 1986b). The value of α is taken from the ratios of covalent radii (r) of atom X relative to Csp^3, thus

$$\alpha_x = r_x / r_{(Csp^3)} - 1$$

The heteroatom weighted ^1k index, $^1k_\alpha$ is then computed, substituting the $1 + \alpha$ value in summing up A and ^1p:

$$^1k_\alpha = (A + \alpha)\ (A + \alpha - 1) / (^1p + \alpha)^2$$

Counts of two and three contiguous bonds in a molecule lead to the formation of second and third order indices.

The zero order kappa index, ^0k, (Kier, 1987) is calculated from the consideration of each atom in isolation.

Physical Nature of Kappa Indices

The physical relevance of kappa values is poorly understood, and as yet they have been little used in QSAR analysis. They were, of course, originally designed to encode the shape of a molecule, and ^0k is thought to include an element of molecular symmetry. More recently Kier (1989) has proposed their use to calculate an index of molecular flexibility (Φ):

$$\Phi = {}^1k_\alpha \cdot {}^2k_\alpha / A$$

Appendix 3. Full Results of the Microtox Test

(The values given are the ml/l (for liquids) and g/l (for solids) concentrations of the chemicals causing a 50% reduction of light output in the Microtox test)

Chemical	Endpoint	Replicates
3-pentanone	5 min	2.6, 2.1, 2.1
	15 min	3.0, 2.5, 2.4
5-nonanone	5 min	0.027, 0.031, 0.033
	15 min	0.039, 0.046, 0.041
3-methyl-2-butanone	5 min	0.08, 0.08, 0.09
	15 min	0.10, 0.090, 0.11
3,3-dimethyl-2-butanone	5 min	0.003, 0.006, 0.0055, 0.0025, 0.0038, 0.031
	15 min	0.0035, 0.0075, 0.0060, 0.0028, 0.0040, 0.0034
2-ethoxyethylacetate	5 min	1.3, 1.0, 1.2
	15 min	1.4, 1.3, 1.4
methyl acetate	5 min	12, 12, 15
	15 min	12, 12, 12
propyl acetate	5 min	0.38, 0.35, 0.35
	15 min	0.45, 0.47, 0.40
butyl acetate	5 min	0.13, 0.10, 0.09, 0.10
	15 min	0.16, 0.13, 0.16, 0.14
hexyl acetate	5 min	0.0095, 0.011, 0.011
	15 min	0.012, 0.015, 0.013
ethyl hexanoate	5 min	0.05, 0.041, 0.052, 0.050
	15 min	0.09, 0.06, 0.059, 0.065
diethyl adipate	5 min	0.035, 0.027, 0.030
	15 min	0.040, 0.030, 0.035
dibutyl adipate	5 min	0.0025, 0.0026, 0.0043
	15 min	0.0024, 0.0026, 0.0047
diethyl sebacate	5 min	0.00058, 0.00062, 0.00075
	15 min	0.00054, 0.00058, 0.00069
dimethyl malonate	5 min	7.0, 9.0, 8.5
	15 min	7.0, 8.5, 7.5
diethyl benzyl malonate	5 min	not toxic at saturation
	15 min	not toxic at saturation
chloroacetonitrile	5 min	0.60, 0.65, 0.60
	15 min	0.18, 0.17, 0.20
malononitrile	5 min	0.40, 0.37, 0.39
	15 min	0.15, 0.155, 0.15
allyl cyanide	5 min	3.5, 3.3, 2.8
	15 min	2.7, 2.8, 2.5
1,4-dicyanobutane	5 min	3.0, 4.5, 1.8, 2.8
	15 min	4.0, 6.0, 2.1, 3.0
1,6-dicyanohexane	5 min	0.017, 0.016, 0.017
	15 min	0.020, 0.019, 0.019
octyl cyanide	5 min	0.0015, 0.00155, 0.0016, 0.0012
	15 min	0.00088, 0.0013, 0.0012, 0.0011
acetone	5 min	21.5, 24.0, 26.5
	15 min	20.5, 24.0, 25.0
toluene	5 min	0.032, 0.031, 0.018
	15 min	0.038, 0.036, 0.021
2-methoxyethylamine	5 min	0.032, 0.035, 0.022, 0.028
	15 min	0.026, 0.031, 0.018, 0.025

1,2-diaminopropane	5 min	0.032, 0.029, 0.028, 0.025
	15 min	0.027, 0.026, 0.025, 0.022
butanal	5 min	0.36, 0.42, 0.29, 0.27
	15 min	0.21, 0.27, 0.23, 0.20
propylamine	5 min	0.021, 0.019, 0.014, 0.016
	15 min	0.013, 0.012, 0.013
2-chloro-4-methylaniline	5 min	0.0046, 0.0041, 0.0054, 0.0044
	15 min	0.0055, 0.0045, 0.0061, 0.0050
octylamine	5 min	0.035, 0.038, 0.035, 0.038
	15 min	0.033, 0.035, 0.032, 0.036
hexanal	5 min	0.042, 0.040, 0.031
	15 min	0.022, 0.028, 0.027
heptylamine	5 min	0.033, 0.035, 0.036
	15 min	0.028, 0.030, 0.035
4-fluoroaniline	5 min	0.092, 0.089, 0.048, 0.060
	15 min	0.095, 0.095, 0.052, 0.072
N,N-diethylaniline	5 min	0.0080, 0.0063, 0.0064
	15 min	0.0096, 0.0075, 0.0076
2-fluorobenzaldehyde	5 min	0.018, 0.018, 0.016, 0.014
	15 min	0.017, 0.019, 0.016, 0.015
2-chloro-6-fluorobenzaldehyde	5 min	0.027, 0.025, 0.030, 0.031
	15 min	0.019, 0.018, 0.024, 0.024
5-bromosalicylaldehyde	5 min	0.0098, 0.0084, 0.0095, 0.0082
	15 min	0.0072, 0.0065, 0.0076, 0.0056
vanillin	5 min	0.070, 0.060, 0.060
	15 min	0.085, 0.060, 0.058
2,4-dichlorobenzaldehyde	5 min	0.0050, 0.0066, 0.0052, 0.0046
	15 min	0.0050, 0.0062, 0.0054, 0.0042
4-chloro-3-nitrotoluene	5 min	0.0043, 0.0043, 0.0032
	15 min	0.0046, 0.0044, 0.0033
1,2,4-trichlorobenzene	5 min	0.00092, 0.0012, 0.0013
	15 min	0.0012, 0.0014, 0.0014
2-chloronitrobenzene	5 min	0.0046, 0.0037, 0.0041
	15 min	0.0046, 0.0039, 0.0046
3-chloronitrobenzene	5 min	0.0115, 0.0115, 0.0092
	15 min	0.012, 0.012, 0.010
2-chloro-4-nitrotoluene	5 min	0.0028, 0.0038, 0.0034
	15 min	0.0038, 0.0042, 0.0036
2-chloro-6-nitrotoluene	5 min	0.00075, 0.00069, 0.00073
	15 min	0.00083, 0.00078, 0.00076
acrolein	5 min	0.00027, 0.00042, 0.00045, 0.00043
	15 min	0.000105, 0.00016, 0.00017 0.00015
biphenyl	5 min	0.0030, 0.0024, 0.0041
	15 min	0.0030, 0.0025, 0.0044
1,3-dichloro-2-propanol	5 min	1.78, 1.30, 1.15
	15 min	1.30, 1.30, 1.15
3-chlorotoluene	5 min	0.0022, 0.0019, 0.0039
	15 min	0.0026, 0.0021, 0.0043
4-chloronitrobenzene	5 min	0.0235, 0.017, 0.0305
	15 min	0.026, 0.019, 0.033